One for All and All for One:
Regression Checks With Many Regressors

Pascal Lavergne

Toulouse School of Economics (GREMAQ)

Valentin Patilea

CREST (ENSAI) and IRMAR (Université Européenne de Bretagne)

Abstract

We develop a novel approach to build checks of parametric regression models when many regressors are present, based on a class of sufficiently rich semiparametric alternatives, namely single-index models. We propose an omnibus test based on the kernel method that performs against a sequence of directional nonparametric alternatives as if there was one regressor only, whatever the number of regressors. This test can be viewed as a smooth version of the integrated conditional moment (ICM) test of Bierens. Qualitative information can be easily incorporated into the procedure to enhance power. In an extensive comparative simulation study, we find that our test is little sensitive to the smoothing parameter and performs well in multidimensional settings. We then apply it to a cross-country growth regression model.

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 et 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 monograph by Hart (1997) nicely 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 wanes when there is more than a couple of regressors. Indeed, as 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 with respect to the sample size.

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 restrictions on the alternative model, but do not yield omnibus tests.

Our goal is thus to devise a powerful regression check that researchers could confidently apply in the presence of many regressors without imposing restrictions on the form of the alternative. It can be viewed as an elaboration of the Integrated Conditional Moment (ICM) test of Bierens (1982), and for this reason we label our test the smooth ICM test. Moreover, compared to existing tests, an appealing feature of our approach is that it allows to incorporate a priori qualitative information in 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 against directional alternatives as if there was one regressor only.

Acknowledging that testing directly against saturated alternatives yield low power, we replace the nonparametric alternative by a *class* of a semiparametric single-index alternatives that is sufficiently rich to allow detection of any nonparametric alternative, thus reducing the dimension of the problem while preserving consistency. Single-index alternatives have been used in previous work on regression checks. Ramsey (1969) first proposed a regression specification error test (RESET) that implicitly consider alternatives depending on a single-index, where the index of interest is taken as the best linear projection of Y on X. Bierens (1982) based the ICM test on correlations between residuals and the set of functions $\psi(X'\beta) = \exp(iX'\beta)$ for any direction β. Further work along these lines include Bierens (1990), Stinchcombe and White (1998), Zhu and Li (1998), Bierens and Ploberger (1997), Escanciano (2006), and Lavergne and Patilea (2008). We detail some of these approaches later on. In this work, we also elaborate on this approach. 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. The null hypothesis then writes

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

As we face the "curse of dimensionality" in estimating the above conditional expectation,

the resulting estimate is imprecise in small and moderate samples, and the related test lacks 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

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

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, the rate of convergence of the test statistic under H_0 equals the rate one would obtain in the one-dimensional case, so that the test behaves against local directional alternatives as if there was one regressor only. We also show that when the regressors are bounded, which is a weak requirement, 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 are always of the same sign, one can choose restrict to the integral over the domain where the corresponding components of β have the same sign.

The paper is organized as follows. Section 2 explains the principle on which our approach relies and compare it to previous approaches. 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 compares our approach to different tests previously proposed in the literature. Section 5 applies our test to a cross-country growth regression as studied by Liu and Stengos (1999). Section 6 suggests directions for future research. The technical proofs are gathered in the Appendix.

2 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. When X is bounded, it is even sufficient to consider *infinitely many* of these conditional expectations. Note that this 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* $X \in \mathbb{R}^q$. *Let* $f_\beta(\cdot)$ *be the density of* $X'\beta$ *and assume that* $\mathbb{E}(Z|x'\beta)f_\beta(x'\beta) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ *for all* $\beta \in \mathbb{S}^q$.
(i) $\mathbb{E}(Z \mid X) = 0$ *is equivalent to*

$$\int_{\mathbb{S}^q} \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right] d\beta = 0 \; . \tag{2.1}$$

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

$$\int_B \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right] d\beta = 0 \tag{2.2}$$

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

Lemma 2.1 can be deduced from Bierens (1982, Theorem 1), see also Escanciano (2006, Lemma 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 $t \in \mathbb{R}$,

$$\psi_\beta(t) := \mathbb{E}\left[\exp\{itX'\beta\}\mathbb{E}(Z \mid X'\beta)\right] = \mathbb{E}\left[\exp\{itX'\beta\}\mathbb{E}(Z \mid X)\right] \; , \tag{2.3}$$

where $i = \sqrt{-1}$. Since for any $\beta \in \mathbb{S}^q$, $\mathbb{E}\left(Z|x'\beta\right) f_\beta(x'\beta) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, Plancherel's theorem, see e.g. Rudin (1987), yields

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

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

If the last integral is 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 (2.2). Since

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

(2.2) 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 (2.1) 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 connected 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 does not contain a half-hypersphere. We also note that our result would adapt when some of the components of X are discrete with a finite support. However we do not pursue this issue further since the dimensionality issue arises in theory only for continuous regressors, namely the rate of convergence of the test statistic is affected by the dimension of X only when X is continuous.

Our approach consists in estimating the quantity in (2.1) and testing whether it is zero. It 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

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

for some probability measure $\mu(\cdot)$ and a well-chosen function $\psi(\cdot)$. Bierens (1982) based his test on estimation of the above quantity with $\psi(\cdot) = \exp(i\cdot)$, Bierens and Ploberger (1997) did the same with $\psi(\cdot) = \exp(\cdot)$. Escanciano (2006) considered a test statistic that estimates

$$\int_{\mathbb{S}^q} \int_{\mathbb{R}^q} \left|\mathbb{E}\left[Z\mathbb{I}\left(X'\beta \leq t'\beta\right)\right]\right|^2\, dF_X(t)\, d\beta \,,$$

which is also zero iff H_0 holds, where $\mathbb{I}(\cdot)$ is the indicator function. From Lemma 2.1, our approach can be similarly viewed as based on the integral of density-weighted expectations, that is

$$\int \left|\mathbb{E}\left[Z\psi_\beta\left(X'\beta\right) f_\beta(X'\beta)\right]\right|^2\, d\beta \,.$$

6

Instead of choosing particular $\psi_\beta(\cdot)$ at the outset, we choose for each β the L^2-function that maximizes the above integrand. We clearly have to normalize the functions $\psi_\beta(\cdot)$ to obtain a unique set of finite solutions. Under our assumptions, a convenient normalization is

$$\mathbb{E}\left[\psi_\beta^2\left(X'\beta\right)f_\beta(X'\beta)\right] = \mathbb{E}\left[\mathbb{E}^2(Z\mid X'\beta)f_\beta(X'\beta)\right] . \tag{2.4}$$

Hence the problem is to find the function that minimizes $\mathbb{E}\left[(Z-\psi\left(X'\beta\right))^2 f_\beta(X'\beta)\right]$, so that $\psi_\beta(X'\beta) = \mathbb{E}(Z\mid X'\beta)$. When X is bounded, we can incorporate some qualitative information by restricting attention to a subset B of $\mathbb{S}^q$ and estimate (2.2) instead of (2.1). We illustrate this possibility later on.

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 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 (2.2), 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

$$Q_n(\beta) = \frac{1}{n(n-1)}\sum_{j\neq i}\widehat{U}_i\widehat{U}_j\frac{1}{h}K_h\left((X_i-X_j)'\beta\right) , \tag{3.1}$$

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 we need here. The quantity in (2.2) is thus

estimated by

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

Let v_n^2 be the variance of $nh^{1/2} I_n$ under H_0, which is strictly positive and finite as shown later. 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. Now

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

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

$$\sup_{1 \leq i \leq n} \left| \frac{\widehat{\sigma}^2(X_i)}{\sigma^2(X_i)} - 1 \right| = o_{\mathbb{P}}(1) , \tag{3.2}$$

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

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

where l 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 (3.2), including in particular the continuity of $\sigma^2(\cdot)$. 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 (3.2) holds.

The use of a nonparametric estimator of the error's variance does not affect our 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 yields

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

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 allow for its use.

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. Lavergne and Patilea (2008) consider a test based on $Q_n(\widehat{\beta}_n)$, with

$$\widehat{\beta}_n = \arg\max_{\mathbb{S}^q} nh^{1/2} Q_n(\beta) - \alpha_n \mathbb{I}\left(\beta \neq \beta^*\right),$$

where β^* represents a favored direction and α_n is a slowly diverging penalty sequence. Their procedure allows to incorporate some information on the preferred single-index alternative, as defined through β^*, but introduces a supplementary user-chosen parameter α_n.

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 the interval $(0, 1)$ through a monotonic transformation.

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 parameters for the technical analysis that follows, i.e. the model is

$$Y = (1, X')\theta + \varepsilon \qquad \mathbb{E}\left[\varepsilon|X\right] = 0\,.$$

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 (2008) 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 does not contain a half-hypersphere, X is assumed to be bounded.

10

(d) Let $\mathbf{Z} = [Z_i, i = 1, \ldots n] = [(1, X_i'), i = 1, \ldots n]$ *be the design matrix. There exists a positive definite matrix A such that $n^{-1}\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \mathbf{A}$. $\theta \in \Theta$, 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) rules out multicollinearity among the regressors and is sufficient for our Lemma 2.1 to hold, so that we adopt it for the sake of simplicity. Note that for X bounded, 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 $X_i's$.*

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, Gonzalez Manteiga, and Presedo Quindimil (1998). Here we use a generalization of this method, the smooth conditional moments bootstrap introduced by Gozalo (1997), which often delivers better small sample results. 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 (3.2), the bootstrap critical value yields a test based on I_n with asymptotic level α conditionally on the original sample.*

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

$$\mathbb{E}[(1, X')\delta(X)] = \mathbf{0} \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(X)] < \infty . \tag{4.1}$$

The first condition ensures that $\delta(\cdot)$ is orthogonal to any linear combination of the regressors. By contrast to what is done in previous literature, we do not impose smoothness restrictions on the function $\delta(\cdot)$. We consider the sequence of local directional alternatives

$$H_{1n}: \ \mathbb{E}\left[Y|X\right] = (1, X')\,\theta_0 + r_n\delta(X), \quad n \geq 1 . \tag{4.2}$$

Such directional alternatives can be detected by our test if $r_n^2 n h^{1/2} \to \infty$, 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, $r_n^2 n h^{q/2} \to \infty$ is needed for consistency. Hence, from a theoretical viewpoint, 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 $r_n^2 n h^{1/2} \to \infty$, the test based on I_n is consistent conditionally on the X_i's against the sequence of alternatives H_{1n} with $\delta(X)$ satisfying (4.1).*

This result allows for a theoretical comparison between different procedures. The ICM test as well as the test proposed by Escanciano (2006) do not involve any smoothing parameter, and thus are consistent against directional alternatives as long as $r_n^2 n \to \infty$. On the other hand, the test proposed by Lavergne and Patilea (2008) is consistent against directional alternatives with $r_n^2 n h^{1/2}/\alpha_n \to \infty$. Hence asymptotic analysis implies a ranking between the different tests. However, nothing ensures that such a ranking will be valid in small or moderate samples. For instance, Lavergne and Patilea (2008) found in their simulation that their test outperforms the ICM and Escanciano's test. Hence we now turn to compare the small sample power of our test with some of its competitors.

4.2 Small sample power

Our simulation study had as a main objective to evaluate the comparative performances of the smooth ICM test (hereafter, SICM) with respect to competing tests based on single indexes. Let us present briefly the different tests we considered. The ICM test (hereafter ICM) is based on the statistic

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

where $\phi(\beta)$ is the standard normal density on $\mathbb{R}^q$, see Bierens (1982). Escanciano (2006) also used this form of ICM 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 should not make any substantial difference. Moreover, as noted by a referee, a Hilbert space approach in asympotic theory could allow for a non-compact support, see e.g. Escanciano (2006). The ICM statistic thus resembles ours, with a kernel depending only on the norm $\|X_i - X_j\|$ but with a fixed bandwidth. Dominguez (2004) shows that the wild bootstrap is valid and preserves asymptotic admissibility of ICM, consequently we used this method to obtain critical values. Escanciano's test is based on the statistic

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

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 based RESET on estimating a third order polynomial regression in the index $X\widehat{\beta}$, where $\widehat{\beta}$ is the OLS estimator of the linear model, and testing for the significance of the non-linear terms. As recommended by Godfrey and Orme (2004), we use restricted residuals and the wild bootstrap in implementing RESET. In practice, we have found that computation of this test is often plagued by imperfect multicollinearity, so the reported results should be viewed with caution. We also considered the test proposed by Lavergne and Patilea (2008), that depends not only on a bandwidth, but also on a penalty sequence α_n as well as a favored direction β^*. Finally, we considered Zheng's test (1996) based on the knowledge

of the true index. Since this test uses information that are not used by the other tests, it constitutes a benchmark against which to judge the performances of each test. We did not directly consider the smoothing test of Zheng (1996) and Li and Wang (1996) because it would involve a comparison between one-dimensional and multi-dimensional smoothing, and it is unclear how this should be done.

A referee suggested to consider the same setup than Zhu (2003) and Escanciano (2006) with multiple regressors and a quadratic alternative. However, the population R^2, that is the part of the variation of the dependent variable explained by the linear projection on the regressors' space, is always greater than 99% in the considered models and this hardly corresponds to any realistic situation. Our setup considers 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 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

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

with $\theta_0 = (0, 0.5, 0.5, 0.5, -1.5)$, and different d and $\delta(\cdot)$. Under the null hypothesis, the population R^2 is 75%. 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 SICM where numerical integration is performed on a grid of 5000 points on an half-hypersphere, and SICM where integration is performed on a grid of 5000 points on the subset B of the hypersphere for which the first three components of β are positive. This corresponds to the knowledge that the marginal effects 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 $h = b \, n^{-2/9}$. This rate is the optimal bandwidth rate derived by Guerre and Lavergne (2002) when the regression function is twice differentiable. As a basis, we chose $b = 1$, but we checked that varying this constant does not affect our qualitative results, see below. The errors' conditional variance was estimated by applying a kernel estimator with normal kernel and bandwidth $cn^{-1/6}$ to the parametric residuals. As a basis, we chose $c = 2$, since it yielded

14

a better behavior for Zheng's test, but our conclusions are not dependent of this choice, see below.

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 $n = 50$ and $d = 12$. Figure 2 compares the power curves of the different tests for $n = 50$, $b = 1$, and $c = 2$. Empirical levels are well approximated by the bootstrap for the smooth tests. Bierens' and Escanciano's tests are under rejecting, with respective levels 2.9 and 1.16. We checked that this phenomenon was not due to too low a number of bootstrap replications and that it was not dependent on the signal-to-noise ratio. However underrejection may be case-dependent, e.g. one referee found accurate levels for a linear model with four regressors but without a constant. In terms of power, SICM outperforms all its competitors but the infeasible Zheng's test in dimension 1. Such a finding is in line with the one of Miles and Mora (2003), who found in simulations that when the regressors' dimension increases non smooth tests such as ICM can be largely outperformed even by multidimensional smooth tests. ICM and Escanciano's test perform almost similarly, and are both more powerful than the RESET test. When incorporating some qualitative information on the first three components, the power of SICM becomes close to the one of the infeasible test Zheng's test.

We then compared ICM and SICM to the test of Lavergne and Patilea (2008, hereafter LP). We chose $\alpha_n = a \log(n)$, where the benchmark $a = 0.2$ is found as suggested in the original paper. We considered two cases. Figure 3 reports the results in which the favored direction corresponds to the true index. For varying a, the power of LP's test is always higher than the one of SICM. We then chose as a favored direction $(2,0,0,1)/\sqrt{3}$, which is orthogonal to the true direction β_0. Figure 4 shows that this changes the ranking. Not only SICM is now more powerful, but LP's test becomes as or less powerful than ICM depending on the choice of α_n.

Figure 5 illustrates that the power of the smooth ICM test varies little when the bandwidth h varies. Figure 6 considers a variation in the bandwidth used for variance estimation, and reaches a similar conclusion. In Figures 7 and 8, we let the sample size varies. For a sample size of 25, the power of all tests is much lower, but the overall

15

comparative behavior of the tests does not change with the exception that Escanciano's test has now almost the same power as RESET. For a sample size of 100, the power curves of all tests come closer, so that it is difficult to rank them. Only SICM on B is clearly more powerful than its competitors, and behaves almost as the infeasible Zheng's test.

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 9 is the analog of Figure 2 for hyperbolic sine alternatives. The RESET is now very powerful, followed closely by SICM on B. Recall however that RESET suffers from imperfect multicollinearity among the terms of the polynomial regression. SICM is more powerful than ICM, followed by Escanciano's test. Figure 10 shows that varying the bandwidth h does not affect much the SICM power. For $n = 100$, the ranking of the different tests is unaffected, as seen in Figure 11.

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 12 shows that for $n = 50$ ICM performs better than Escanciano's test, but SICM outperforms them both. Restricting integration on B further improves its power. Surprisingly, RESET does even better than SICM. Figure 13 shows that for a large bandwidth SICM and ICM have similar performances, as could be expected.

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 14 for $n = 50$ shows the previous qualitative findings still hold. Figure 15 illustrates that they still hold when the bandwidth varies.

Two main features emerge from our simulations. First, when many regressors are present, the smooth ICM test is powerful in most cases, likely 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, as our sine alternatives, the smooth ICM test can do a better job in small samples. Clearly, such findings are not accounted for by asymptotic results.

5 Empirical application

A vast literature has used cross-country regressions to search for linkages between long-run average growth rates and a variety of economic policy, political and institutional factors. Such work uses the classical Solow model as a basic framework, see the contributions of Barro (1991) and Mankiw et al. (1992). In this application, we used data from King and Levine (1993), and specifically pooled cross-country data averaged over the 1960s, 1970s and 1980s for 86 countries as investigated by Liu and Stengos (1999). We consider the linear regression of GDP, the growth rate of the gross domestic product, on $\ln(POP)$, the natural logarithm of population growth (adjusted for depreciation rate and technological change), $\ln(INV)$, the natural logarithm of the share of output allocated to investment, $\ln(GDP60)$, the natural logarithm of the initial level of GDP in 1960, and $\ln(SEC)$, the natural logarithm of human capital as measured by the enrolment rate in secondary schools. Fixed-effects time dummies are also included. Theory and empirics suggest that the GDP growth is decreasing in population growth, and increasing in both human capital and investment. There is more uncertainty about the relationship with the initial level of GDP. The estimated regression is

$$
\begin{aligned}
\widehat{GDP} = \quad & 0.0299 && -0.0324\ln(POP) && +0.0286\ln(INV) && -0.0037\ln(GDP60) \\
& (0.0281) && (0.0111) && (0.0055) && (0.0026) \\
& && +0.0037\ln(SEC) && +\text{Time fixed effects} \\
& && (0.0021)
\end{aligned}
$$

with clustered standard errors into parentheses. This corresponds to our expectations, though the coefficient of $\ln(GDP60)$ has a large standard error. To check nonlinearities in $\ln(GDP60)$ and $\ln(SEC)$, one can introduce quadratic then cubic terms, as in Liu and Stengos (1999), but the coefficients of these polynomial terms are insignificant (jointly for each variable), with p-values greater than 0.4 in any case (likely due to imperfect multicollinearity). Hence simple parametric methods do not point in the direction of a more sophisticated model. However, Liu and Stengos (1999) showed using semiparametric methods that there exist nonlinearities in the (logarithms of) initial level of GDP and human capital. Specifically, growth is as expected increasing in the logarithm of human capital, but at a decreasing rate, while growth is first increasing then decreasing in initial GDP.

17

We applied different tests considered in the simulations with X the vector of the four continuous variables $\ln(POP)$, $\ln(INV)$, $\ln(GDP60)$, and $\ln(SEC)$. For the smooth ICM test, we also considered a version that incorporates our expectations about the relationships between growth and population, human capital and investment. That is, we integrate over the domain of the hypersphere for which the respective signs of the components of β match our expectations. Before implementing the tests, we center and rescale the matrix $\mathbf{X}$. Other details of the computation are the same as in the simulations.

The p-value for the ICM and Escanciano's test were 0.0854 and 0.1106 respectively. So in this application, neither would reject the linear model at a 5% significance level. Table 1 reports the results of the different nonparametric tests. By contrast, the p-value for the smooth ICM test varies between 1 and 2%. The smooth ICM restricted on B has p-value zero irrespective of chosen bandwidth. These results indicates that the smooth ICM test and its variant unambiguously detect some unaccounted nonlinearities in the growth regression.

6   Concluding remarks

Our smooth ICM test allows to check the functional form a regression model with many regressors. The principle of our test is to replace one conditional expectation given all the explanatory variables by all conditional expectations given one single linear index. The same idea can be used to test general conditional moment restrictions, as studied for instance by Lewbel (1995), Donald, Imbens and Newey (2003), and Delgado, Dominguez, and Lavergne (2006). This line of research is currently under investigation.

Acknowledgements

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 6.1 *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) Let $U_n = \frac{1}{n(n-1)} \sum_{j \neq i} \delta(X_i)\delta(X_j)h^{-1}\mathbb{E}_B\left[K_h\left((X_i - X_j)'\beta\right)\right]$. If $\mathbb{E}\delta^4(X) < \infty$ and $nh \to \infty$, then $U_n - \mathbb{E}(U_n) = o_{\mathbb{P}}(1)$.

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 integral is strictly positive and bounded by

$$
\int_{\mathbb{R}} \int_{\mathbb{S}^q} \left|\widehat{\delta f}(t\beta)\right|^2 \, d\beta\, dt = 2\pi \int_{\mathbb{S}^q} \mathbb{E}\left[\mathbb{E}^2(\delta(X)|X'\beta)f_\beta(X'\beta)\right]\, d\beta < \infty\,.
$$

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

and the above expectation converges to a finite limit from Part (i). ∎

Let $\mathbf{W}$ be the matrix with generic element $\mathbb{E}_B\left[K_h\left((X_i - X_j)'\beta\right)\right]\mathbb{I}(i \neq j)/(hn(n-1))$ and define its spectral radius as $\mathrm{Sp}(\mathbf{W}) = \sup_{u \neq 0} \|\mathbf{W}u\|/\|u\|$.

Lemma 6.2 *Under Assumptions D(c) and K, (i)* $\text{Sp}(\mathbf{W}) = O_{\mathbb{P}}(n^{-1})$ *and (ii)* $n^2 h \|\mathbf{W}\|^2$ *has a strictly positive limit, where* $\|\mathbf{W}\|$ *denotes the Euclidean matrix norm.*

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

$$
\|\mathbf{W}u\|^2 = \sum_{i=1}^{n} \left(\sum_{j=1, j \neq i}^{n} w_{ij} u_j \right)^2 \leq \sum_{i=1}^{n} \left(\sum_{j=1, j \neq i}^{n} w_{ij} \right) \sum_{j=1, j \neq i}^{n} w_{ij} u_j^2
$$

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

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

$$
\mathbb{P} \left[\left(\frac{(nh^2)^\alpha}{\ln n} \right)^{1/2} \max_{1 \leq i \leq n} \frac{1}{(n-1)h} \left| \sum_{j \neq i} \{ \mathbb{E}_B K_h \left((X_i - X_j)' \beta \right) - \mathbb{E} \left[\mathbb{E}_B K_h \left((X_i - X_j)' \beta \right) | X_i \right] \} \right| \geq t \right]
$$

$$
\leq \sum_{1 \leq i \leq n} \mathbb{E} \left[\mathbb{P} \left[\left| \frac{1}{(n-1)} \sum_{j \neq i} \{ \mathbb{E}_B K_h \left((X_i - X_j)' \beta \right) \right. \right. \right.
$$

$$
\left. \left. \left. - \mathbb{E} \left[\mathbb{E}_B K_h \left((X_i - X_j)' \beta \right) | X_i \right] \} \right| \geq t h \left(\frac{\ln n}{(nh^2)^\alpha} \right)^{1/2} | X_i \right] \right]
$$

$$
\leq 2n \exp \left(-\frac{t^2}{2} \frac{(nh^2)(\ln n)}{C((nh^2)^\alpha + th(nh^2)^{\alpha/2}(\ln n)^{1/2})} \right) \leq 2 \exp \left[\ln n - \frac{t^2}{C'} (\ln n)(nh^2)^{1-\alpha} \right] \to 0,
$$

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 $n^2 h \|\mathbf{W}\|^2 = \frac{1}{(n-1)^2} \sum_{i \neq j} h^{-1} \mathbb{E}_B^2 K_h \left((X_i - X_j)' \beta \right)$. Hoeffding's (1963) inequality for U-statistics yields for any $t > 0$

$$
\mathbb{P} \left[\frac{1}{n(n-1)h} \left| \sum_{j \neq i} \{ \mathbb{E}_B^2 K_h \left((X_i - X_j)' \beta \right) - \mathbb{E} \left[\mathbb{E}_B^2 K_h \left((X_i - X_j)' \beta \right) \right] \} \right| \geq t \right]
$$

$$
= \mathbb{P} \left[\left| \frac{1}{n(n-1)} \sum_{j \neq i} \{ \mathbb{E}_B^2 K_h \left((X_i - X_j)' \beta \right) - \mathbb{E} \left[\mathbb{E}_B^2 K_h \left((X_i - X_j)' \beta \right) \right] \} \right| \geq th \right]
$$

$$
\leq 2 \exp \left(-\frac{t^2 (nh^2)}{C} \right) \to 0,
$$

by Assumption K(b). First note that

$$0 < \mathbb{E}\left[h^{-1}\mathbb{E}_B^2 K_h\left((X_i - X_j)'\beta\right)\right] < C\mathbb{E}\left[h^{-1}\mathbb{E}_B K_h\left((X_i - X_j)'\beta\right)\right] < \infty.$$

Moreover

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

For $B = \mathbb{S}^q$, we have by Assumption K(a)

$$
\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{S}^q}\int_{\mathbb{S}^q}|\widehat{K}(u)|\left|\widehat{f}\left(t\beta + h^{-1}u\alpha\right)\right|^2 dt\, du\, d\beta\, d\alpha \\
&= \int_{\mathbb{R}\times\mathbb{S}^q}\left|\widehat{f}\left(t\beta\right)\right|^2 dt\, d\beta \int_{\mathbb{S}^q} d\alpha \int_{\mathbb{R}}|\widehat{K}(u)|\, du < \infty\,.
\end{aligned}
\tag{6.1}
$$

Now

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

From the uniform continuity of $\widehat{K}(\cdot)$ and Equation (6.1), the right-hand side can be rendered arbitrarily small by choosing M small enough then letting h tend to zero. Therefore the limit of $\mathbb{E}\left[h^{-1}\mathbb{E}_B^2 K_h\left((X_i - X_j)'\beta\right)\right]$ is, by Lebesgue dominated convergence theorem and Assumption K(a),

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

which is positive by Assumption K(a). ∎

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

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

Under Assumption D, $\widehat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$. Hence $I_{2n} \leq \mathrm{Sp}(\mathbf{W})\|\mathbf{Z}(\widehat{\theta}_n - \theta_0)\|^2 = O_{\mathbb{P}}(n^{-1})$ by Lemma 6.2(i). Let $\mathbb{E}_n$ denote the conditional expectation given the X_i, Z_k be any column of $\mathbf{Z}$, $k = 1, \ldots d + 1$, and $\overline{Z}_k = Z_k'\mathbf{W}$. Then Marcinkiewicz-Zygmund's and Minkowski's inequalities imply that there is some C independent of n such that

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

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 $\|\mathbf{W}\|^{-1}\mathrm{Sp}(\mathbf{W}) = o_{\mathbb{P}}(1)$. Lemma 6.2 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 $I_{3n} = r_n\delta'\mathbf{W}\mathbf{Z}(\widehat{\theta}_n - \theta_0)$, $I_{4n} = r_n\delta'\mathbf{W}\varepsilon$, and $I_{5n} = r_n^2\delta'\mathbf{W}\delta$. By Assumption D(c) and Lemma 6.2(ii), $v_n^2 \leq \overline{\sigma}^4 n^2 h\|\mathbf{W}\|^2 = O_{\mathbb{P}}(1)$. 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 $|u'\mathbf{W}v| \leq \|u\|\|v\|\mathrm{Sp}(\mathbf{W})$, $r_n^{-1}I_{3n} \leq \|\delta\|\|\mathbf{Z}(\widehat{\theta}_n - \theta_0)\|\mathrm{Sp}(\mathbf{W}) = O_{\mathbb{P}}(n^{-1/2})$. Also $I_{4n} = O_{\mathbb{P}}(r_n n^{-1/2})$ by the same arguments used for dealing with I_{1n}. Lemma 6.1(ii) yields $I_{5n} = r_n^2 C + o_{\mathbb{P}}(r_n^2)$ with $C > 0$. Collecting results, it follows that $nh^{1/2}I_n = nh^{1/2}r_n^2 C + o_{\mathbb{P}}(r_n^2 nh^{1/2})$. Deduce from $\widehat{v}_n^2/v_n^2 = O_{\mathbb{P}}(1)$ and $r_n^2 nh^{1/2} \to \infty$ 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 (447)**, 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.

BARRO, R.J. (1991). Economic growth in a cross-section of countries. *Quart. J. Econ.* **106 (2)**, 407–443.

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 W. PLOBERGER (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65 (5)**, 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.

DELGADO, M.A., M.A. DOMINGUEZ, AND P. LAVERGNE (2006). Consistent tests of conditional moment restrictions. *Ann. Econom. Statist.* **81 (1)**, 33–67.

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

DONALD, S.G., G.W. IMBENS, AND W.K. NEWEY (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics* **117 (1)**, 55–93.

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.

ESCANCIANO, J.C. (2006). Goodness-of-fit tests for linear and nonlinear rime series models. *J. Amer. Statist. Assoc.* **101 (474)** 531–541.

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.

GODFREY, L.G., AND C.D. ORME (2004). Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients. *Economics Letters* **82 (2)**, 282-287.

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

GUERRE, E., AND P. LAVERGNE (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.

KING, R.G., AND R. LEVINE (1993). Finance and growth: Schumpeter might be right. *Quart. J. Econ.* **108 (3)**, 717–737.

LAVERGNE, Q., AND V. PATILEA (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143 (1)**, 103–122.

LEWBEL, A. (1995). Consistent nonparametric hypothesis tests with an application to Slutsky symmetry. *J. Econometrics* **67 (1)**, 379–401.

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

LIU, Z., AND T. STENGOS (1999). Non-linearities in cross-country growth regressions: a semi-parametric approach. *J. Appl. Econometrics* **14 (5)**, 527–538.

MANKIW, N.G., D. ROMER, AND D.N. WEIL (1992). A contribution to the empirics of economic growth. *Quart. J. Econ.* **107 (2)**, 407–437.

MILES, D., AND MORA, J. (2003). On the performance of nonparametric specification tests in regression models. *Comput. Statist. Data Anal.* **42 (3)**, 477–490.

RAMSEY, J.B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **31 (2)**, 511-534.

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 QUINDIMIL (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 n=50

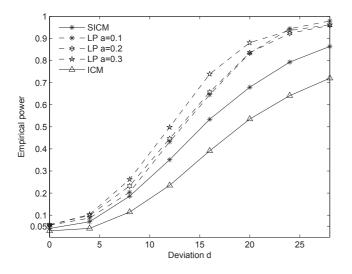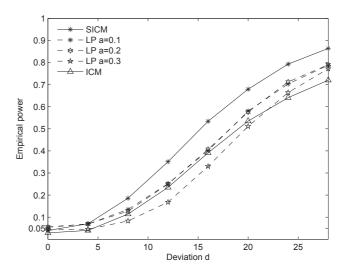Figure 3: Quadratic alternative n=50 — $\beta^* = \beta_0$

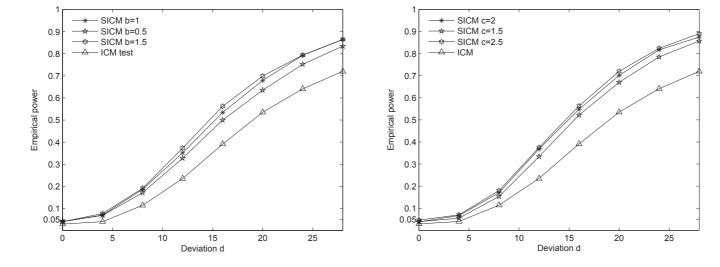Figure 4: Quadratic alternative n=50 — $\beta^* \perp \beta_0$

Figure 5: Quadratic alternative n=50 — varying h

Figure 6: Quadratic alternative n=50 — varying l
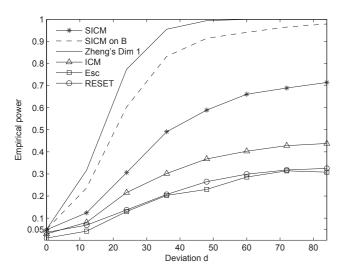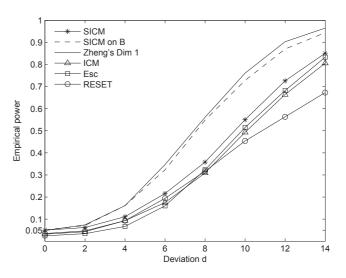

Figure 7: Quadratic alternative n=25
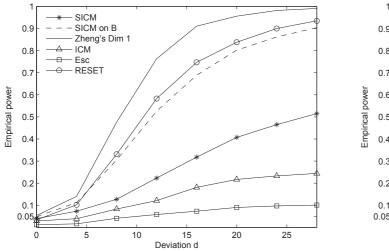

Figure 8: Quadratic alternative n=100

Figure 9: Sinh alternative n=50
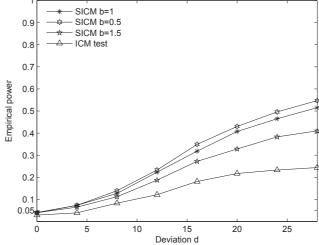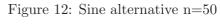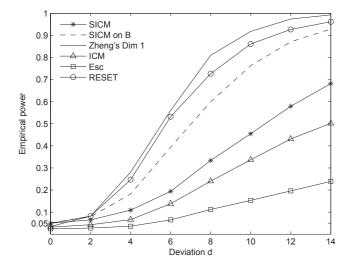

Figure 10: Sinh alternative n=50 — varying h
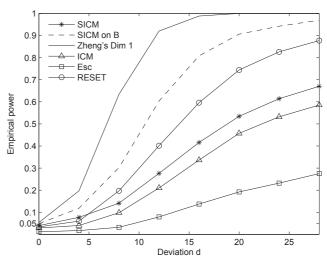

Figure 11: Sinh alternative n=100

Figure 12: Sine alternative n=50

Table 1:

Bootstrap p-values (in %)for growth regression

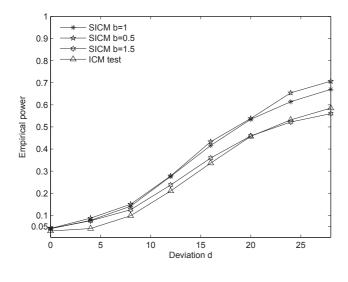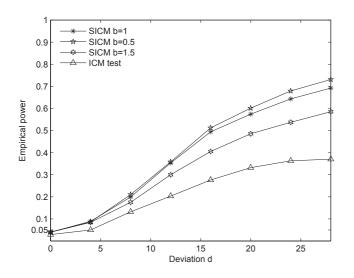| b	0.5	0.75	1	1.25	1.5
SICM	2.01	1.51	1.01	1.01	2.01
SICM on B	0	0	0	0	0

Figure 13: Sine alternative n=50 — varying h

Figure 14: Two-indices alternative n=50

Figure 15: Two-indices alternative n=50 — varying h