Econometric Theory, **22**, 2006, 1030–1051. Printed in the United States of America.
DOI: 10.1017/S0266466606060506

A CONSISTENT DIAGNOSTIC TEST FOR REGRESSION MODELS USING PROJECTIONS

J. Carlos Escanciano
Universidad de Navarra

This paper proposes a consistent test for the goodness-of-fit of parametric regression models that overcomes two important problems of the existing tests, namely, the poor empirical power and size performance of the tests due to the curse of dimensionality and the subjective choice of parameters such as bandwidths, kernels, and integrating measures. We overcome these problems by using a residual marked empirical process based on projections (RMPP). We study the asymptotic null distribution of the test statistic, and we show that our test is able to detect local alternatives converging to the null at the parametric rate. It turns out that the asymptotic null distribution of the test statistic depends on the data generating process, and so a bootstrap procedure is considered. Our bootstrap test is robust to higher order dependence, in particular to conditional heteroskedasticity. For completeness, we propose a new minimum distance estimator constructed through the same RMPP as in the testing procedure. Therefore, the new estimator inherits all the good properties of the new test. We establish the consistency and asymptotic normality of the new minimum distance estimator. Finally, we present some Monte Carlo evidence that our testing procedure can play a valuable role in econometric regression modeling.

1. INTRODUCTION

The purpose of the present paper is to develop a consistent, powerful, and simple diagnostic test for testing the adequacy of a parametric regression model with the property of being free of any user-chosen parameter (e.g., bandwidth) and, at the same time, being suitable for cases in which the covariate is of high or moderate finite dimension. Most consistent tests proposed in the literature give misleading results for this latter empirically relevant case. This problem is intrinsic and is often referred to as the "curse of dimensionality" in the regression literature; see Section 7.1 of Fan and Gijbels (1996) for some discussion on this problem. More precisely, let $(Y, X')'$ be a random vector in

The author thanks Carlos Velasco and Miguel A. Delgado for useful comments. The paper has also benefited from the comments of two referees and the co-editor. This research was funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001. Address correspondence to Juan Carlos Escanciano, Facultad de Economicas, Universidad de Navarra, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain; e-mail: jescanci@unav.es.

  0266-4666/06 $12.00

a $(d+1)$-dimensional Euclidean space, where Y represents the real-valued dependent (or response) variable, X is the d-dimensional explanatory variable, $d \in \mathbb{N}$, and A' denotes the matrix transpose of A. Under $E|Y| < \infty$, it is well known that the regression function $m(x) = E[Y|X = x]$ is well defined. If in addition $E|Y|^2 < \infty$, then $m(X)$ represents almost surely (a.s.) the "best" prediction of Y given X, in a mean square sense. Then, it is common in regression modeling to consider the following tautological expression:

$$Y = m(X) + \varepsilon,$$

where $\varepsilon = Y - E[Y|X]$ is, by construction, the unpredictable part (in mean) of Y given X.

Much of the existing literature is concerned with parametric modeling in that m is assumed to belong to a given parametric family $\mathcal{M} = \{f(\cdot,\theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ and, by analogy, one considers the following parametric regression model:

$$Y = f(X,\theta) + e(\theta), \tag{1}$$

with $f(X,\theta)$ a parametric specification for the regression function $m(X)$ and with $e(\theta)$ a random variable (r.v.), disturbance of the model. Parametric regression models continue to be attractive to practitioners because these models have the appealing property that the parameter θ together with the functional form $f(\cdot,\cdot)$ describes, in a very concise way, the relation between the response Y and the explanatory variable X. Because we do not know in advance the true regression model, to prevent wrong conclusions, every statistical inference that is based on model f should be accompanied by a proper model check. As a matter of fact, a correct specification of m is important in model-based economic decisions and/or to interpret parameters correctly.

Note that $m \in \mathcal{M}$ is tantamount to

$$E[e(\theta_0)|X] = 0 \quad \text{a.s.,} \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p. \tag{2}$$

There is a vast amount of literature on testing consistently the correct specification of a parametric regression model. Although the idea of the proposed consistent tests is similar in all cases, namely, comparing a parametric and a (semi-) nonparametric estimation of a functional of the conditional mean in (2), they can be divided into two classes of tests. The first class of tests uses nonparametric smoothing estimators of $E[e(\theta_0)|X]$. We call this approach the "local approach"; see Eubank and Spiegelman (1990), Eubank and Hart (1992), Wooldridge (1992), Yatchew (1992), Gozalo (1993), Härdle and Mammen (1993), Horowitz and Härdle (1994), Hong and White (1995), Zheng (1996), Li (1999), Horowitz and Spokoiny (2001), Koul and Ni (2004), and Guerre and Lavergne (2005) for some examples. A methodology related to the local approach is that of empirical likelihood procedures as proposed in Chen, Härdle, and Li (2003) and Tripathi and Kitamura (2003). The local approach requires smoothing of

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

the data in addition to the estimation of the finite-dimensional parameter vector and leads to less precise fits. Tests based on the local approach have standard asymptotic null distributions, but their finite-sample distributions depend on the choice of a bandwidth (or similar) of the nonparametric estimator, which affects the inference procedures.

The second class of tests avoids smoothing estimation by means of reducing the conditional mean independence in (2) to an infinite (but parametric) number of unconditional orthogonality restrictions, i.e.,

$$E[e(\theta_0)|X] = 0 \quad \text{a.s.} \Leftrightarrow E[e(\theta_0)w(X,x)] = 0, \quad \forall x \in \Pi, \tag{3}$$

where Π is a properly chosen space and the parametric family $w(\cdot,x)$ is such that the equivalence (3) holds; see Bierens and Ploberger (1997), Stinchcombe and White (1998), and Escanciano (2006) for primitive conditions on the family $w(\cdot,x)$ to satisfy this equivalence. We call the approach based on (3) the "integrated approach" because it uses the integrated (cumulative) measures of dependence $E[e(\theta_0)w(X,x)]$. In the literature the most frequently used weighting functions have been the exponential function, e.g., $w(X,x) = \exp(ix'X)$ in Bierens (1982, 1990), where $i = \sqrt{-1}$ denotes the imaginary unit and the indicator function $w(X,x) = 1(X \leq x)$; see, e.g., Stute (1997), Koul and Stute (1999), Whang (2000), and Li, Hsiao, and Zinn (2003), among many others. Different families w deliver different power properties of the integrated-approach-based tests. Most tests based on the integrated approach have nonstandard asymptotic null distributions, but they can be well approximated by bootstrap methods; see, e.g., Stute, Gonzalez-Manteiga, and Presedo-Quindimil (1998).

An important problem with the local approach arises when the dimension of the explanatory variable X is high or even moderate. The sparseness of the data in high-dimensional spaces leads most local-based test statistics to suffer a considerable bias, even for large sample sizes. This is an important practical limitation for most tests considered in the literature, because it is not uncommon in econometric modeling to have high-order models. Some statistical theories have been developed to overcome this problem; cf. generalized linear models (GLM) (see, e.g., McCullagh and Nelder, 1989) or single-index models (see, e.g., Powell, Stock, and Stoker, 1989). However, these theories are semiparametric and, therefore, need smoothing techniques. In addition, they do not cover all possible models.

Here, we propose a new consistent test within the integrated framework that compares very well to indicator- and exponential-based tests. The new test is simple to compute, does not need user-chosen parameters or high-dimensional numerical integration, is robust to higher order dependence (in particular to conditional heteroskedasticity), and presents excellent empirical power properties in finite samples; see Section 4. Furthermore, our test procedure provides a formalization of some well-known traditional exploratory tools based on residual-fitted values plots.

The organization of the paper is as follows. In Section 2 we define the residual marked process based on projections (RMPP) as the basis for our test statistic. In Section 3 we study the asymptotic null distribution and the behavior against Pitman‘s local alternatives of the new test statistic. For completeness of exposition, we consider in this section a new minimum distance estimator for the regression parameter based on the RMPP, and we show its consistency and asymptotic normality under similar assumptions as in the testing procedure. Also, because the asymptotic null distribution depends on the data generating process (DGP), a bootstrap procedure to approximate the asymptotic critical values of the test statistic is considered. In Section 4 we conduct a simulation exercise comparing the new proposed test with some competing tests considered in the literature. This Monte Carlo experiment shows that our new test can play a valuable role in parametric regression modeling. Proofs of the main results are deferred to Appendix A. Appendix B contains a simple algorithm to compute the new test statistic.

2. THE RESIDUAL MARKED PROCESS BASED ON PROJECTIONS

Let $\{Z_i = (Y_i, X_i')'\}_{i=1}^n$ be a sequence of independent and identically distributed (i.i.d.) $(d+1)$-dimensional r.v.s defined on the probability space $(\Omega, \mathcal{A}, P)$ and with the same distribution as $Z = (Y, X')'$, with $0 < E|Y| < \infty$. The main goal in this paper is to test the null hypothesis (2), i.e.,

$$H_0 : E[Y|X] = f(X, \theta_0) \quad \text{a.s.} \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p,$$

against the alternative

$$H_A : P(E[Y|X] \neq f(X, \theta)) > 0, \quad \text{for all } \theta \in \Theta \subset \mathbb{R}^p.$$

As argued before, one way to characterize H_0 is by the infinite number of parametric unconditional moment restrictions

$$E[e(\theta_0) w(X, x)] = 0, \quad \forall x \in \Pi, \tag{4}$$

where the parametric family $w(\cdot, x)$ is such that the equivalence in (3) holds. Examples of such families are $w(X,x) = 1(X \leq x)$, $w(X,x) = \exp(ix'X)$, $w(X,x) = \sin(x'X)$, and $w(X,x) = 1/(1 + \exp(c - x'X))$ with $c \neq 0$; see the aforementioned references for many other families.

In view of a sample $\{Z_i\}_{i=1}^n$, define the marked empirical process

$$R_{n,w}(x, \theta) = n^{-1/2} \sum_{i=1}^{n} e_i(\theta) w(X_i, x). \tag{5}$$

Define also $R_{n,w}^1(\cdot) \equiv R_{n,w}(\cdot, \theta_n)$, where θ_n is an $\sqrt{n}$-consistent estimator of θ_0. The marks in $R_{n,w}^1$ are given by the classical residuals; therefore, we call $R_{n,w}^1$ a residual marked empirical process.

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

Because of the equivalence (3), it is natural to base the tests on a distance from $R^1_{n,w}$ to zero, i.e., on a norm $\Gamma(R^1_{n,w})$, say. The most used norms are the Cramér–von Mises (CvM) and Kolmogorov–Smirnov (KS) functionals:

$$
\begin{aligned}
CvM_{n,w} &= \int_{\Pi} |R^1_{n,w}(x)|^2 \Psi(dx),\\
KS_{n,w} &= \sup_{x\in\Pi} |R^1_{n,w}(x)|,
\end{aligned}
\tag{6}
$$

respectively, where $\Psi(x)$ is an integrating function satisfying some mild conditions; see A4 in Section 3. Other functionals are possible. Then, tests in the integrated approach reject the null hypothesis (2) for "large" values of $\Gamma(R^1_{n,w})$.

The first consistent integrated test proposed in the literature was that of Bierens (1982) based on the exponential weighting family, i.e., using the residual marked process

$$
R^1_{n,\exp}(x) = n^{-1/2} \sum_{i=1}^{n} e_i(\theta_n) \exp(ix'\Phi(X_i)),
$$

where $\Phi(\cdot)$ is a bounded one-to-one Borel measurable mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$. Bierens (1982) considered a CvM norm with integrating measures $\Psi(dx) = \Upsilon(x)\,dx$, with $\Upsilon(x) = 1(x \in \Pi_{l=1}^d [-\varepsilon_l, \varepsilon_l])$, where $\varepsilon_l > 0$, $l = 1, \dots, d$, are arbitrarily chosen numbers (Bierens, 1982, p. 109), and $\Upsilon(x)$ equals a d-variate normal density function (Bierens, 1982, p. 111).

On the other hand, Stute (1997) used the indicator family $w(X,x) = 1(X \le x)$ in the residual marked process. The main advantage of the indicator weighting function over the exponential function is that it avoids the choice of an arbitrary integrating function Ψ, because in the indicator case this is given by the natural empirical distribution function of $\{X_i\}_{i=1}^n$. On the contrary, the indicator weight has the drawback of being more sensitive to the dimension d than the exponential weight, which is based on one-dimensional projections (see Escanciano, 2006).

In this paper we propose a test based on a new family $\{w, \Psi\}$ of weighting and integrating functions that possesses the good properties of the exponential- and indicator-based tests and at the same time prevents their deficiencies. The new test avoids the arbitrary choice of the integrating function or numerical integration in high-dimensional spaces and is less sensitive to the dimension d than indicator-based tests because it is based on one-dimensional projections. The CvM test based on this new family presents an excellent performance in finite samples and is very simple to compute. In addition, the new family w formalizes some traditional model diagnostic tools based on residual-fitted values plots for linear models.

Our first aim is to avoid the problem of the curse of dimensionality. The following result can be viewed as a particularization of the Cramér–Wold principle to our main concern, the goodness-of-fit of the regression function. The term $|A|$ denotes the Euclidean norm of A.

LEMMA 1. *A necessary and sufficient condition for (2) to hold is that for any vector $\beta \in \mathbb{R}^d$ with $|\beta| = 1$,*

$$E[e(\theta_0)|\beta' X] = 0 \quad \text{a.s.,} \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p.$$

Lemma 1 yields that consistent tests for H_0 can be based on one-dimensional projections. In particular, we have the characterization of the null hypothesis H_0:

$$H_0 \Leftrightarrow E[e(\theta_0) 1(\beta' X \le u)] = 0$$
$$\text{almost everywhere (a.e.) } (\beta, u) \in \Pi, \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \tag{7}$$

where from now on $\Pi = \mathbb{S}_d \times [-\infty, \infty]$ is the nuisance parameter space with $\mathbb{S}_d$ the unit ball in $\mathbb{R}^d$, i.e., $\mathbb{S}_d = \{\beta \in \mathbb{R}^d : |\beta| = 1\}$. Therefore, the test we consider here rejects the null hypothesis for "large" values of the standardized sample analogue of $E[e(\theta_0) 1(\beta' X \le u)]$.

An approach related to ours is that of Stute and Zhu (2002), who considered the weighting family $\{1(\beta_0' X \le u)\}$ for model checks of GLM in an i.i.d. framework. However, note that they fix the direction to β_0, the direction involved in the GLM, and so their approach is clearly different from that considered here, because we consider all the directions β in $\mathbb{S}_d$ simultaneously. As a consequence, our test will be consistent against all alternatives, whereas in our present framework the Stute and Zhu (2002) test is only consistent against alternatives satisfying that $E[e(\theta_*) 1(\beta_*' X \le u)] \ne 0$ in a set with positive Lebesgue measure in $\mathbb{R}$, where θ_* and β_* are the probabilistic limits under the alternative of the estimators of θ_0 and β_0, respectively.

The family $1(\beta' X \le u)$ yields the RMPP

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} e_i(\theta_n) 1(\beta' X_i \le u).$$

The marks of R_n^1 are given by the classical residuals and the "jumps" by the projected regressors. Note that for a fixed direction β, R_n^1 is uniquely determined by the residuals and the projected variables $\{\beta' X_i\}_{i=1}^n$ and vice versa. As in the usual residual-regressors plot, we can plot the path of R_n^1 for different directions β as an exploratory diagnostic tool. In particular, in the linear model, the plot of the path of $R_n^1(\beta_n, u)$, with β_n the least squares estimator, resembles the usual residual-fitted values plot. Therefore, tests based on $R_n^1(\beta_n, u)$ provide a formalization of such traditional well-known exploratory tools.

To measure the distance from R_n^1 to zero a norm has to be chosen. From computational considerations a CvM norm is very convenient in our context. Two facts motivate our choice of the integrating measure in the CvM norm.

First, note that once the direction β is fixed, u lives in the projected regressor variable's space, and second, in principle, all the directions are equally important; cf. Lemma 1. To define our CvM test we need some notation. Let $F_{n,\beta}(u)$ be the empirical distribution function of the projected regressors $\{\beta' X_i\}_{i=1}^n$ and $d\beta$ the uniform density on the unit sphere. Also let $F_\beta(u)$ be the true cumulative probability distribution function (c.d.f.) of $\beta' X$. Then, we define the new CvM test as

$$PCvM_n = \int_\Pi (R_n^1(\beta,u))^2 F_{n,\beta}(du)\, d\beta. \tag{8}$$

Therefore, we reject the null hypothesis H_0 for large values of $PCvM_n$. See Appendix B for a simple algorithm to compute $PCvM_n$ from a given data set $\{Z_i\}_{i=1}^n$. The next section justifies inference for $PCvM_n$ based on the asymptotic theory.

Our test statistic $PCvM_n$ avoids the deficiencies of the Bierens (1982) and Stute (1997) tests, namely, the arbitrary choice of the integrating function or numerical integration in high-dimensional spaces and the low power performance when the dimension d is large, respectively. However, it is worthwhile to mention that our test is not necessarily better than the Bierens (1982) and Stute (1997) tests. In fact, using the results of Bierens and Ploberger (1997) it can be shown that all these tests are asymptotically admissible, and therefore none of them is strictly better than the others uniformly over the space of alternatives. However, in our simulations that follow we show that for the alternatives considered our test is the best or comparable to the best test. A simple intuition as to why our test performs so well with the alternatives considered is as follows. Under the alternative it can be shown that, uniformly in $x \in \Pi$,

$$n^{-1/2} R_{n,w}^1(x) \xrightarrow{P^*} E[e(\theta_1) w(X,x)],$$

where θ_1 is the probabilistic limit of θ_n under the alternative H_A. On the other hand, under the normalization $E[m^2(X,\theta_1)] = 1$, where $m(\cdot,\theta_1) = E[e(\theta_1)| X = \cdot]$, it holds that the optimization problem

$$\max_{w,\, E[w^2(I_{t-1})]=1} |E[e_t(\theta_1) w(I_{t-1})]|^2$$

attains its optimum at $w^*(\cdot) = m(\cdot,\theta_1)$. Therefore, as $w(\cdot,\cdot)$ is closer to $m(\cdot,\cdot)$, the test based on w is expected to have better power properties. It seems that for the models considered in Section 4 $m(\cdot,\theta_1)$ can be "well approximated" by our weight function $1(\beta' X \le u)$, and this might explain the good power properties of our test procedure.

During the revision process of the paper one of the referees suggested a modification of our test that might have better finite-sample performance. Based on the inequality

$$\int_{-\infty}^{\infty} (E[\varepsilon 1(\beta' X \le u)])^2 F_\beta(du) \le (E[\varepsilon \sqrt{1 - F_\beta(\beta' X)}])^2,$$

which follows from simple algebra, the modified test statistic is

$$\int_{\mathbb{S}_d} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(\theta_n) \sqrt{1 - F_{n,\beta}(\beta' X_i)} \right)^2 d\beta.$$

However, contrary to $PCvM_n$ the latter test statistic involves numerical integration and is much more difficult to compute. Therefore, we do not study this modified test statistic further in the paper. On the contrary, the next section studies the asymptotic distribution theory for $PCvM_n$.

3. ASYMPTOTIC THEORY

Now, we establish the limit distribution of R_n^1 under the null hypothesis H_0. For the asymptotic theory, note that R_n^1 can be viewed as a mapping from $(\Omega, \mathcal{A}, P)$ with values in $\ell^\infty(\Pi)$, the space of all real-valued functions that are uniformly bounded on Π. Let $\Rightarrow$ denote weak convergence on $\ell^\infty(\Pi)$ and $\xrightarrow{P^*}$ denote convergence in outer probability; see Definitions 1.3.3 and 1.9.1, respectively, in van der Vaart and Wellner (1996). Also, $\xrightarrow{d}$ stands for convergence in distribution of real r.v.s. To derive asymptotic results we consider the following assumptions. First, let us denote by $F_Y(\cdot)$ and $F_X(\cdot)$ the marginal c.d.f. of Y and X, respectively. Also let $\Psi_p(\cdot)$ be the product measure of $F_\beta(\cdot)$ and the uniform distribution on $\mathbb{S}_d$, i.e., $\Psi_p(d\beta, du) = F_\beta(du)\, d\beta$. In the discussion that follows C is a generic constant that may change from one expression to another.

Assumption A1.

A1(a) $\{Z_i = (Y_i, X_i')'\}_{i=1}^n$ is a sequence of i.i.d. random vectors with $0 < E|Y_i| < \infty$.
A1(b) $E|\varepsilon|^2 < C$.

Assumption A2. $f(\cdot,\theta)$ is twice continuously differentiable in a neighborhood Θ_0 of θ_0, $\Theta_0 \subset \Theta$. The score $g(X,\theta) = (\partial/\partial\theta') f(X,\theta)$ verifies that there exists a $F_X(\cdot)$-integrable function $M(\cdot)$ with $sup_{\theta\in\Theta_0} |g(\cdot,\theta)| \le \mathrm{M}(\cdot)$.

Assumption A3.

A3(a) The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ. There exists a $\theta_1 \in \Theta$ such that $|\theta_n - \theta_1| = o_P(1)$.
A3(b) The estimator θ_n satisfies the following asymptotic expansion under H_0:

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(Y_i, X_i, \theta_0) + o_P(1),$$

where $l(\cdot)$ is such that $E[l(Y, X, \theta_0)] = 0$ and $L(\theta_0) = E[l(Y, X, \theta_0) l'(Y, X, \theta_0)]$ exists and is positive definite.

Assumption A4. $\Psi_p(\cdot)$ is absolutely continuous with respect to Lebesgue measure on Π.

Assumptions A1 and A2 are standard in the model checks literature; see, e.g., Bierens (1990) and Stute (1997). Assumption A3 is satisfied, e.g., for the nonlinear least squares estimator and (under further regularity assumptions) its robust modifications; see, e.g., Chapters 5 and 7 in Koul (2002). Note that A3(a) and A3(b) imply that $\theta_0 = \theta_1$ under the null H_0, but they are not necessarily equal under the alternative. We shall show subsequently that A3 is also satisfied for a new minimum distance estimator. Assumption A4 is only necessary for consistency of the test.

Under A1 and (2), using a classical central limit theorem (CLT) for i.i.d. sequences, we have that the finite-dimensional distributions of R_n, where R_n is the process defined in (5) with $\theta = \theta_0$ and $w(X, x) = 1(\beta' X \leq u)$, converge to those of a multivariate normal distribution with a zero mean vector and variance-covariance matrix given by the covariance function

$$K(x_1, x_2) = E[\varepsilon^2 1(\beta_1' X \leq u_1) 1(\beta_2' X \leq u_2)], \tag{9}$$

where $x_1 = (\beta_1', u_1)'$ and $x_2 = (\beta_2', u_2)'$. The next result is an extension of this convergence to weak convergence in the space $\ell^\infty(\Pi)$. Throughout the paper $x = (\beta', u)'$ will denote the nuisance parameter, and we interchange the notation x and $(\beta', u)'$ whenever this does not create confusion.

THEOREM 1. *Under the null hypothesis H_0 and Assumption A1*

$$R_n \Rightarrow R_\infty,$$

where $R_\infty(\cdot)$ is a Gaussian process with zero mean and covariance function given by (9).

In practice θ_0 is unknown and has to be estimated from a sample $\{Z_i\}_{i=1}^n$ by an estimator θ_n, say. The next result shows the effect of the parameter uncertainty on the asymptotic null distribution of R_n^1. To this end, let us define the function $G(x, \theta_0) = G(x) = E[g(X, \theta_0) 1(\beta' X \leq u)]$ and let V be a normal random vector with zero mean and variance-covariance matrix given by $L(\theta_0)$ as defined in A3(b).

THEOREM 2. *Under the null hypothesis H_0 and Assumptions A1–A3*

$$R_n^1(\cdot) \Rightarrow R_\infty(\cdot) - G'(\cdot) V \equiv R_\infty^1(\cdot),$$

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

where R_∞ *is the same process as in Theorem 1 and*

$$\mathrm{cov}(R_\infty(x), V) = E[\varepsilon l(Y, X, \theta_0) 1(\beta' X \le u)].$$

Theorem 2 and the continuous mapping theorem (CMT) (see, e.g., van der Vaart and Wellner, 1996, Thm. 1.3.6) yield the asymptotic null distribution of the functional $PCvM_n$.

COROLLARY 1. *Under the assumptions of Theorem 2, for any continuous functional (with respect to the supremum norm)* $\Gamma(\cdot)$,

$$\Gamma(R_n^1) \xrightarrow{d} \Gamma(R_{\infty,w}^1).$$

Furthermore,

$$PCvM_n \xrightarrow{d} PCVM_\infty = \int_\Pi (R_\infty^1(\beta, u))^2 \Psi_p(d\beta, du).$$

Note that the integrating measure in $PCvM_n$ is a random measure, but Corollary 1 shows that the asymptotic theory is not affected by this fact. Also note that the asymptotic null distribution of $PCvM_n$ depends in a complex way on the DGP and the specification under the null, and so critical values have to be tabulated for each model and each DGP, making the application of these asymptotic results difficult in practice. To overcome this problem we approximate the asymptotic null distribution of continuous functionals of R_n^1 by a bootstrap procedure given subsequently.

In Assumption A3 we require that the estimator of θ_0 admits an asymptotic linear representation. For completeness of the presentation we give some mild sufficient conditions under which a minimum distance estimator (see Koul, 2002, Ch. 5 and references therein), is asymptotically linear. Motivated from Lemma 1, we have that under the null

$$\theta_0 = \arg\min_{\theta \in \Theta} \int_\Pi |E[e(\theta) 1(\beta' X \le u)]|^2 \Psi_p(d\beta, du) \tag{10}$$

and θ_0 is the unique value that satisfies (10). Then, we propose estimating θ_0 by the sample analogue of (10), i.e.,

$$\theta_n = \arg\min_{\theta \in \Theta} \int_\Pi n^{-1} |R_n^1(\beta, u, \theta)|^2 F_{n,\beta}(du)\, d\beta. \tag{11}$$

This estimator is a minimum distance estimator and extends in some sense the generalized method of moments (GMM) estimator, frequently used in econometric and statistical applications. This kind of generalization of GMM was considered first in Carrasco and Florens (2000) for univariate problems. Recently, and for $w(X, x) = 1(X \le x)$, Dominguez and Lobato (2004) have considered an estimator similar to (11) for a conditional moment restriction under time series. Also using this principle, Koul and Ni (2004) have proposed a minimum distance estimation for θ_0 using an L_2-distance similar to that used in Härdle and Mammen (1993) in the "local approach." Our estimator θ_n has the advantage of being free of any user-chosen parameter (bandwidth, kernel, or integrating measure) and is expected to be more robust to the problem of the curse of dimensionality than the estimating procedures based on $1(X \leq x)$ or local approaches. Now, we shall show that θ_n in (11) satisfies Assumption A3. The following matrices are involved in the asymptotic variance-covariance matrix of the estimator:

$$C = \int_{\Pi} G(\beta,u)G'(\beta,u)\Psi_p(d\beta,du),$$

$$D = \int_{\Pi\times\Pi} G(x)G'(x)K(x,y)\Psi_p(dx)\Psi_p(dy).$$

For the consistency and asymptotic normality of the estimator we need an additional assumption.

Assumption A1$'$. The regression function $f(\cdot,\theta)$ satisfies that there exists an $F_X(\cdot)$-integrable function $K_f(\cdot)$ with $\sup_{\theta\in\Theta}|f(\cdot,\theta)| \leq K_f(\cdot)$.

THEOREM 3. *Under H_0 and Assumptions A1, A2, and A1$'$*

(i) the estimator given in (11) is consistent, i.e., $\theta_n \to \theta_0$ a.s.;
(ii) if in addition, the matrix C is nonsingular, then

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N(0, C^{-1}DC^{-1}).$$

From Theorem 3 we have immediately the asymptotic linear expansion required in A3(b):

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l(Y_i, X_i, \theta_0) + o_P(1),$$

where now

$$l(Y_i,X_i,\theta_0) = -C^{-1}\{Y_i - f(X_i,\theta_0)\}\int_{\Pi} G(\beta,u)1(\beta' X_i \leq u)\Psi_p(d\beta,du).$$

Note that in general the estimator given in (11) is not asymptotically efficient. An asymptotically efficient estimator based on the same minimum distance principle can be constructed following the ideas of Carrasco and Florens (2000). This optimal estimator will require the choice of a regularization parameter needed to invert a covariance operator; see Carrasco and Florens (2000) for more details.

Now we study the asymptotic distribution of R_n^1 under a sequence of local alternatives converging to null at a parametric rate $n^{-1/2}$. We consider the local alternatives

$$H_{A,n}: Y_{i,n} = f(X_i, \theta_0) + \frac{a(X_i)}{n^{1/2}} + \varepsilon_i, \quad \text{a.s.}, \qquad 1 \leq i \leq n, \tag{12}$$

where the random variable $a(X)$ is F_X-integrable with zero mean and satisfies $P(a(X) = 0) < 1$. To derive the next result we need the following assumption.

Assumption A3′. The estimator θ_n satisfies the following asymptotic expansion under $H_{A,n}$:

$$\sqrt{n}(\theta_n - \theta_0) = \xi_a + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(Y_i, X_i, \theta_0) + o_P(1),$$

where the function $l(\cdot)$ is as in Assumption A3 and ξ_a is a vector in $\mathbb{R}^p$.

Remark 1. It is not difficult to show that θ_n in (11) satisfies Assumption A3′ under Assumptions A1, A2, and A1′ with

$$\xi_a = C^{-1} \int_{\Pi} E[a(X)1(\beta' X \leq u)] G(\beta, u) \Psi_p(d\beta, du).$$

THEOREM 4. *Under the local alternatives (12) and Assumptions A1, A2, and A3′*

$$R_n^1 \Rightarrow R_\infty^1 + D_a,$$

where R_∞^1 is the process defined in Theorem 2 and the function $D_a(\cdot)$ is the determinist function

$$D_a(\beta, u) = E[a(X)1(\beta' X \leq u)] - G'(\beta, u)\xi_a.$$

For some estimators, D_a has an intuitive geometric interpretation. For instance, for the new minimum distance estimator (11) the shift function is given by

$$D_a(\beta, u) = E[a(X)1(\beta' X \leq u)] \\ \qquad - G'(\beta, u) C^{-1} \int_{\Pi} E[a(X)1(\beta' X \leq u)] G(\beta, u) \Psi_p(d\beta, du)$$

and represents the orthogonal projection in $L_2(\Pi, \Psi_p)$, the Hilbert space of all real-valued and Ψ_p-square-integrable functions on Π, of $E[a(X)1(\beta' X \leq u)]$ parallel to $G(\beta, u)$. The next corollary is a consequence of the CMT and Theorem 4.

COROLLARY 2. *Under the local alternatives (12) and Assumptions A1, A2, and A3′, for any continuous functional* $\Gamma(\cdot)$

$$\Gamma(R_n^1) \xrightarrow{d} \Gamma(R_\infty^1 + D_a).$$

Furthermore,

$$\int_\Pi |R_n^1(\beta,u)|^2 F_{n,\beta}(du)\, d\beta \xrightarrow{d} \int_\Pi |R_\infty^1(\beta,u) + D_a(\beta,u)|^2 \Psi_p(d\beta,du).$$

Note that because of Lemma 1, we have that

$$D_a = 0 \Leftrightarrow a(X) = \xi_a' g(X,\theta_0) \quad \text{a.s.}$$

Therefore, from this result it is not difficult to show that the test based on $PCvM_n$ is able to detect asymptotically any local alternative $a(\cdot)$ that is not parallel to $g(\cdot,\theta_0)$. This result is not attainable for tests based on the local approach, e.g., the Härdle and Mammen (1993) test.

We have seen before that the asymptotic null distribution of continuous functionals of R_n^1 depends in a complicated way on the DGP and the specification under the null. Therefore, critical values for the test statistics cannot be tabulated for general cases. Here we propose to implement the test with the assistance of a bootstrap procedure. Resampling methods have been extensively used in the model checks literature of regression models; see, e.g., Stute et al. (1998) or more recently Li et al. (2003). It is shown in these papers that the most relevant bootstrap method for regression problems is the wild bootstrap (WB) introduced in Wu (1986). We approximate the asymptotic null distribution of R_n^1 by that of

$$R_n^*(x) = n^{-1/2} \sum_{i=1}^n e_i^*(\theta_n^*) 1(\beta' X_i \le u) \qquad x = (\beta',u)' \in \Pi,$$

where the sequence $\{e_i^*(\theta_n^*)\}_{i=1}^n$ are the fixed design wild bootstrap (FDWB) residuals computed from $e_t^*(\theta_n^*) = Y_i^* - f(X_i,\theta_n^*)$, where $Y_i^* = f(X_i,\theta_n) + e_i(\theta_n)V_i$, θ_n^* is the bootstrap estimator calculated from the data $\{(Y_i^*,X_i')'\}_{i=1}^n$, and $\{V_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with zero mean, unit variance, and bounded support and also independent of the sequence $\{Z_i\}_{i=1}^n$. Examples of $\{V_i\}_{i=1}^n$ sequences are i.i.d. Bernoulli variates with

$$P(V_i = a_1) = p_1 \qquad P(V_i = a_2) = 1 - p_1, \tag{13}$$

where $a_1 = 0.5(1-\sqrt{5})$, $a_2 = 0.5(1+\sqrt{5})$, and $p_1 = (1+\sqrt{5})/2\sqrt{5}$, used in, e.g., Li et al. (2003). For other sequences see Mammen (1993). The reader is referred to Stute et al. (1998) for the theoretical justification of this bootstrap approximation and the assumptions needed. The results of these authors jointly with those proved here ensure that the proposed bootstrap test has a correct

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

asymptotic level, is consistent, and is able to detect alternatives tending to the null at the parametric rate $n^{-1/2}$. The next section shows that this bootstrap procedure provides good approximations in finite samples.

4. MONTE CARLO EVIDENCE

In this section we compare the new CvM test with some competing integrated-approach-based tests proposed in the literature. This study complements others considered in the literature; see, e.g., Miles and Mora (2003). We briefly describe our simulation setup. We denote by $PCvM_n$ the new CvM test defined in (8). For the explicit computation of $PCvM_n$ see Appendix B.

Bierens (1982, p. 111) proposed the CvM test statistic based on the exponential weight function $w(X,x) = \exp(ix'X)$ and the d-variate normal density function as the integration function, i.e.,

$$CvM_{n,\exp} = n^{-1} \sum_{i=1}^{n} \sum_{s=1}^{n} e_i(\theta_n) e_s(\theta_n) \exp\left(-\frac{1}{2}|X_i - X_s|^2\right).$$

We also consider here the CvM and KS statistics defined in Stute (1997), which are given by

$$CvM_n = \frac{1}{n^2} \sum_{j=1}^{n} \left[\sum_{i=1}^{n} e_i(\theta_n) 1(X_i \le X_j) \right]^2$$

and

$$KS_n = \max_{1 \le j \le n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(\theta_n) 1(X_i \le X_j) \right|,$$

respectively. Note that CvM_n and $PCvM_n$ are the same test statistics when $d = 1$, by definition.

Recently, Stute and Zhu (2002) have considered an innovation process transformation of $R_n^1(\beta_n, u)$ for testing the correct specification of GLM models, where β_n is a suitable estimator of the GLM parameter, say, β_0. More concretely, their test statistic is the CvM test

$$SZ_n = \frac{1}{\psi_{n,\beta_n}^2(x_0)} \int_{-\infty}^{x_0} |T_n R_n^1(\beta_n, u)|^2 \sigma_{n,\beta_n}^2(u) F_{n,\beta_n}(du),$$

where

$$T_n f(u) = f(u) - \int_{-\infty}^{u} a'_{n,\beta_n}(v) A_n^{-1}(v) \int_{v}^{\infty} a_{n,\beta_n}(y) \sigma_{n,\beta_n}^{-2}(y) f(dy) F_{n,\beta_n}(dv),$$

$$A_n(u) = \int_{u}^{\infty} a_{n,\beta_n}(v) a'_{n,\beta_n}(v) \sigma_{n,\beta_n}^{-2}(v) F_{n,\beta_n}(dv),$$

$a_{n,\beta_n}(u)$ and $\sigma^2_{n,\beta_n}(u)$ are Nadaraya–Watson estimators of $a_{\beta_0}(u) = E[g(X,\theta_0)/\beta_0' X = u]$ and $\sigma^2_{\beta_0}(u) = E[\varepsilon^2 | \beta_0' X = u]$, respectively, $\psi_{n,\beta_n}(u) = n^{-1}\sum_{i=1}^n e_i^2(\theta_n)1(\beta_n' X_i \le u)$, and x_0 is the 99% quantile of F_{n,β_n}. Under the correct specification of the GLM and some additional assumptions

$$SZ_n \xrightarrow{d} \int_0^1 B^2(u)\,du,$$

where $B(\cdot)$ denotes a standard Brownian motion on $[0,1]$; see Stute and Zhu (2002) for further details. For the nonparametric estimators we have chosen a Gaussian kernel with bandwidth $h = 0.5n^{-1/2}$ as in Stute and Zhu (2002).

We consider the same FDWB for the version of the exponential Bierens test and for the Stute (1997) test as for our CvM test $PCvM_n$. For SZ_n we consider empirical critical values based on 10,000 simulations on the first null model in each block of models. In the discussion that follows, $\varepsilon_i \sim$ iid $N(0,1)$ and $\nu_i \sim$ iid $\exp(1)$ are standard Gaussian and centered exponential noises, respectively. We consider in the simulations two blocks of models. In the first block the null model is

$$Y_i = a + bX_{1i} + cX_{2i} + \varepsilon_i,$$

where $X_{1i} = (W_i + W_{1i})/2$ and $X_{2i} = (W_i + W_{2i})/2$, and W_i, W_{1i}, and W_{2i} are i.i.d. $U[0,2\pi]$, independent of ε_i, $1 \le i \le n$. We examine the adequacy of this model under the following DGPs:

1. DGP1: $Y_i = 1 + X_{1i} + X_{2i} + \varepsilon_i \equiv X_i'\alpha_0 + \varepsilon_i$.
2. DGP1-EXP: $Y_i = 1 + X_{1i} + X_{2i} + \nu_i = X_i'\alpha_0 + \nu_i$.
3. DGP2: $Y_i = X_i'\alpha_0 + 0.1(W_{1i} - \pi)(W_{2i} - \pi) + \varepsilon_i$.
4. DGP3: $Y_i = X_i'\alpha_0 + X_i'\alpha_0 \exp\{-0.01(X_i'\alpha_0)^2\} + \varepsilon_i$.
5. DGP4: $Y_i = X_i'\alpha_0 + \cos(0.6\pi X_i'\alpha_0) + \varepsilon_i$.

DGP1 and DGP2 are considered in Hong and White (1995). DGP3 here is similar to their DGP3; see also Koul and Stute (1999). DGP4 is similar to that considered in Eubank and Hart (1992). DGP1-EXP is considered here to show the robustness of the tests against fatter tailed error distributions. For the first block of models we consider a sample size of $n = 50$, 100, and 300. The number of Monte Carlo experiments is 1,000, and the number of bootstrap replications is $B = 500$. For the bootstrap approximation we employ the sequence $\{V_i\}_{i=1}^n$ of i.i.d. Bernoulli variates given in (13). We estimate the null model by the usual least squares estimator. The nominal levels are 10%, 5%, and 1%.

In Table 1 we show the empirical rejection probabilities (RP) associated with models DGP1 and DGP1-EXP. The empirical levels of the test statistics are close to the nominal level, even for sample sizes as small as 50. The empirical levels for DGP1-EXP are less accurate than for DGP1 but are reasonable, showing that the tests are robust to fat-tailed error distributions.

TABLE 1. Empirical size of tests

	$n = 50$			$n = 100$			$n = 300$		
DGP1	10%	5%	1%	10%	5%	1%	10%	5%	1%
$PCvM_n$	9.3	5.2	0.8	10.8	5.7	1.1	10.1	5.7	1.0
$CvM_{n,\exp}$	9.5	4.8	0.8	9.8	5.5	1.0	10.5	5.3	1.2
CvM_n	11.0	5.3	0.8	10.8	5.1	1.3	9.8	5.0	1.1
KS_n	11.5	6.0	1.3	12.1	6.3	1.5	10.8	5.9	1.0
SZ_n	10.3	6.2	0.9	9.5	4.5	0.7	11.2	5.0	0.8
DGP1-EXP									
$PCvM_n$	10.1	5.1	0.7	8.6	3.7	0.5	9.0	4.3	0.9
$CvM_{n,\exp}$	11.5	5.8	0.8	9.4	4.2	0.7	8.3	4.4	0.6
CvM_n	9.4	4.7	0.7	9.0	3.7	0.4	8.9	4.2	0.9
KS_n	11.5	5.4	0.8	9.0	3.7	0.5	9.2	4.4	1.2
SZ_n	9.1	4.7	1.4	10.1	4.3	2.0	10.3	5.4	1.4

In Table 2 we report the empirical power against the DGP2. It increases with the sample size n for all test statistics, as expected. It is shown that the new CvM test $PCvM_n$ has the best empirical power in all cases. The empirical power for $CvM_{n,\exp}$ is reasonable and greater than or equal to CvM_n and KS_n for $n = 50$, but better for $n = 100$ and 300. The Stute and Zhu (2002) test, SZ_n, is the worst against this alternative. The rejection probabilities of $PCvM_n$ are comparable to the best test in Hong and White (1995) against this alternative. In Table 3 we show the RP for DGP3. For this alternative SZ_n and our test statistic, $PCvM_n$, have generally the best empirical powers, SZ_n performing slightly better than $PCvM_n$. Bierens' test $CVM_{n,\exp}$ has good power properties for this alternative. Stute's test CvM_n performs similarly to $CVM_{n,\exp}$, whereas KS_n presents the worst results, with a moderate power. For DGP4 (Table 4), $PCvM_n$ and $CVM_{n,\exp}$ have excellent empirical powers. Stute's tests, CvM_n and KS_n, and the Stute and Zhu (2002) test, SZ_n, have low power against this "high-frequency" alternative.

TABLE 2. Empirical power of tests

	$n = 50$			$n = 100$			$n = 300$		
DGP2	10%	5%	1%	10%	5%	1%	10%	5%	1%
$PCvM_n$	23.3	13.0	3.0	43.2	28.7	7.0	91.3	83.6	53.7
$CvM_{n,\exp}$	21.1	11.5	2.9	39.2	26.1	5.9	89.2	79.4	47.8
CvM_n	20.7	11.1	2.6	33.3	21.7	7.0	79.3	65.2	32.2
KS_n	18.4	11.5	2.5	29.3	18.4	5.3	62.5	47.7	23.0
SZ_n	13.5	5.2	1.7	18.8	12.4	3.5	34.2	24.5	10.7

Table 3. Empirical power of tests

	$n=50$			$n=100$			$n=300$		
DGP3	10%	5%	1%	10%	5%	1%	10%	5%	1%
$PCvM_n$	72.7	61.8	32.6	94.8	92.0	77.5	100.0	100.0	100.0
$CvM_{n,\exp}$	68.4	56.6	27.3	93.8	89.8	71.5	100.0	100.0	100.0
CvM_n	66.7	52.0	26.9	93.5	90.6	72.3	100.0	100.0	100.0
KS_n	43.0	27.1	8.2	80.1	68.9	37.3	100.0	99.9	98.5
SZ_n	72.4	56.4	35.1	97.1	93.9	82.9	100.0	100.0	100.0

The second block of models is taken from Zhu (2003). The null model is

$$Y_i = X_i'\gamma_0 + \varepsilon_i,$$

whereas the DGPs considered are

$$Y_i = X_i'\gamma_0 + b(X_i'\beta_0)^2 + \varepsilon_i,$$

where X_i' is a random d-dimensional covariate with i.i.d. $U[0,2\pi]$ marginal components, $d=3$ and 6. When $d=3$, $\gamma_0=(1,1,2)'$ and $\beta_0=(2,1,1)'$, and when $d=6$, $\gamma_0=(1,2,3,4,5,6)'$ and $\beta_0=(6,5,4,3,2,1)'$. Furthermore, set $b = 0.01, 0.02,\dots,0.1$ when $d=3$ and $b=0.001, 0.002,\dots,0.01$ when $d=6$. This experiment provides us with evidence of the power performance of the tests under local alternatives ($b=0$ corresponds to the null hypothesis). The sample size is $n=25$; the rest of the Monte Carlo parameters are as before.

We show the RP for these models in Figure 1. We see that in both cases, $d=3$ and 6, our new test statistic $PCvM_n$ and SZ_n have the best empirical powers for all values of b. None of them is superior to the other for all values of b and for both models. For $d=3$, SZ_n performs slightly better than $PCvM_n$. They are followed by $CvM_{n,\exp}$. For $d=6$, $PCvM_n$ has the best power for $b \le 0.006$, whereas SZ_n is the best for $b > 0.006$; $CvM_{n,\exp}$, CvM_n, and KS_n have very low empirical power against this alternative.

Table 4. Empirical power of tests

	$n=50$			$n=100$			$n=300$		
DGP4	10%	5%	1%	10%	5%	1%	10%	5%	1%
$PCvM_n$	24.1	13.6	2.5	48.3	29.7	6.9	99.9	98.9	71.1
$CvM_{n,\exp}$	24.6	13.3	2.7	51.2	29.3	8.1	99.8	97.9	76.5
CvM_n	11.1	5.5	1.0	14.8	8.1	2.0	41.7	25.3	5.1
KS_n	9.6	4.8	0.4	15.7	8.5	2.0	39.5	25.4	9.2
SZ_n	12.5	5.4	1.1	16.6	9.4	1.8	33.6	16.5	3.8

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

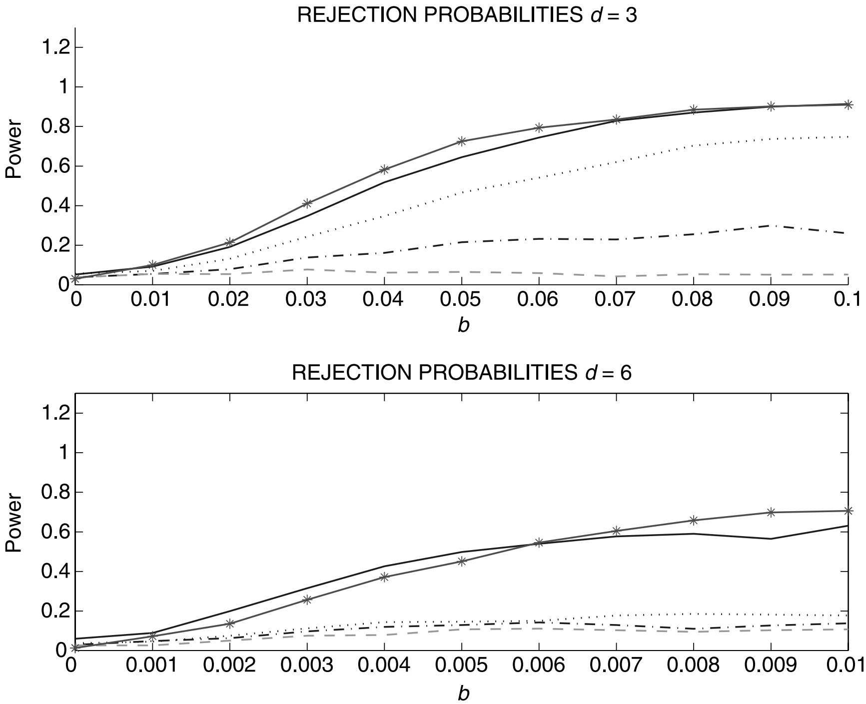

FIGURE 1. Rejection probabilities plots for $d = 3$ and 6. The solid, solid-star, dot, dash, and dash-dot lines are, respectively, for the empirical power of $PCvM_n$, SZ_n, $CvM_{n,\exp}$, CvM_n, and KS_n.

Summarizing, these two Monte Carlo experiments show that our test possesses an excellent power performance in finite samples for the alternatives considered. In all cases, our test has the best empirical power or it is comparable to the best test among the tests proposed by Bierens (1982), Stute (1997), and Stute and Zhu (2002). In our Monte Carlo experiments we have focused on the integrated-approach-based tests. Miles and Mora (2003) have compared through simulations some local-based and integrated-based tests. These authors conclude that for one-dimensional regressors, the integrated-approach-based tests perform slightly better than the smoothing-based ones, especially Bierens' statistic. When the number of regressors is greater than one, some of the smoothing tests considered by these authors perform better. Therefore, it is important to compare our new test with the smoothing-based tests considered by these authors, especially for the case of multivariate regressors. This study is beyond the scope of this paper and is deferred for future research. Our test has the advantage that no bandwidth selection is required, though its implementation requires the use of a bootstrap procedure. Our Monte Carlo experiments show that our test should be considered a reasonably competent test to the best local-approach-based test and a valuable diagnostic procedure for regression modeling.

REFERENCES

Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Ecomometrica* 58, 1443–1458.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment test. *Econometrica* 65, 1129–1151.

Carrasco, M. & J.P. Florens (2000) Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797–834.

Chang, N.M. (1990) Weak convergence of a self-consistent estimator of a survival function with doubly censored data. *Annals of Statistics* 18, 391–404.

Chen, S.X., W. Härdle, & M. Li (2003) An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society, Series B* 65, 663–678.

Dominguez, M. & I. Lobato (2004) Consistent estimation of models defined by conditional moment restrictions. *Econometrica* 72, 1601–1615.

Escanciano, J.C. (2006) Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association* 101, 531–541.

Eubank, R. & J. Hart (1992) Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics* 20, 1412–1425.

Eubank, R. & S. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* 85, 387–392.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451–477.

Guerre, E. & P. Lavergne (2005) Rate-optimal data-driven specification testing for regression models. *Annals of Statistics* 33, 840–870.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1974.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Horowitz, J.L. & W. Härdle (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.

Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.

Jennrich, R.I. (1969) Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.

Koul, H.L. (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. Lecture Notes in Statistics, vol. 166. Springer-Verlag.

Koul, H.L. & P. Ni (2004) Minimum distance regression model checking. *Journal of Statistical Planning and Inference* 119, 109–144.

Koul, H.L. & W. Stute (1999) Nonparametric model checks for time series. *Annals of Statistics* 27, 204–236.

Li, Q. (1999) Consistent model specification test for time series econometric models. *Journal of Econometrics* 92, 101–147.

Li, Q., C. Hsiao, & J. Zinn (2003) Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.

Mammen, E. (1993) Bootstrap and wild bootstrap for high-dimensional linear models. *Annals of Statistics* 21, 255–285.

McCullagh, P. & J. Nelder (1989) *Generalized Linear Models*. Monographs on Statistics and Applied Probability 37. Chapman and Hall.

Miles, D. & J. Mora (2003) On the performance of nonparametric specification tests in regression models. *Computational Statistics & Data Analysis* 42, 477–490.

Powell, J.L., J.M. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

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

Stute, W., W. Gonzalez-Manteiga, & M. Presedo-Quindimil (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* 93, 141–149.

Stute, W. & L.X. Zhu (2002) Model checks for generalized linear models. *Scandinavian Journal of Statistics* 29, 535–545.

Tripathi, G. & Y. Kitamura (2003) Testing conditional moment restrictions. *Annals of Statistics* 31, 2059–2095.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer-Verlag.

Whang, Y.-J. (2000) Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics* 98, 27–46.

Wolfowitz, J. (1954) Generalization of the theorem of Glivenko-Cantelli. *Annals of Mathematical Statistics* 25, 131–138.

Wooldridge, J. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* 8, 452–475.

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

Yatchew, A.J. (1992) Nonparametric regression tests based on least squares. *Econometric Theory* 8, 435–451.

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

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

APPENDIX A: Proofs

Proof of Lemma 1. This follows easily from Part I of Theorem 1 in Bierens (1982). ■

Proof of Theorem 1. By a classical CLT we can show that the finite-dimensional distributions of R_n converge to those of the Gaussian process R_∞. The asymptotic equicontinuity of R_n follows by a direct application of Theorem 2.5.2 in van der Vaart and Wellner (1996); see also their Problem 14 on p. 152. ■

Proof of Theorem 2. Applying the classical mean value theorem argument we have

$$R_n^1(x) = R_n(x) - n^{-1/2}\sum_{i=1}^{n}\{f(X_i,\theta_n) - f(X_i,\theta_0)\}1(\beta' X_i \le u)$$

$$= R_n(x) - I - II - III,$$

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

where

$$I = n^{1/2}(\theta_n - \theta_0)\,\frac{1}{n}\sum_{i=1}^{n}\{g(X_i,\tilde{\theta}_n) - g(X_i,\theta_0)\}1(\beta' X_i \le u),$$

$$II = n^{1/2}(\theta_n - \theta_0)\,\frac{1}{n}\sum_{i=1}^{n}[g(X_i,\theta_0)1(\beta' X_i \le u) - G(x,\theta_0)],$$

and

$$III = n^{1/2}(\theta_n - \theta_0)G(x,\theta_0)$$

and where $\tilde{\theta}_n$ satisfies $|\tilde{\theta}_n - \theta_0| \le |\theta_n - \theta_0|$ a.s. By Assumptions A1–A3, the generalization by Wolfowitz (1954) of the Glivenko–Cantelli theorem, and the uniform law of large numbers (ULLN) of Jennrich (1969), it is easy to show that $I = o_P(1)$ and $II = o_P(1)$ uniformly in $x \in \Pi$. Thus the theorem follows from Theorem 1 and Assumption A3. ■

Proof of Corollary 1. For a nonrandom continuous functional, the result follows from the CMT and Theorem 2. For $PCvM_n$ the result follows because under the conditions of Theorem 2 we have that R_n^1 is asymptotically tight and hence Lemma 3.1 in Chang (1990) applies. ■

Proof of Theorem 3. The proof follows exactly the same steps as the proof of Theorems 1 and 2 in Dominguez and Lobato (2004), and thus it is omitted. ■

Proof of Theorem 4. Under the local alternatives (12) write

$$R_n^1(x) = n^{-1/2}\sum_{i=1}^{n}\left\{f(X_i,\theta_0) + \frac{a(X_i)}{n^{1/2}} + \varepsilon_i - f(X_i,\theta_n)\right\}1(\beta' X_i \le u)$$

$$= R_n(x) + A_1 + A_2, \qquad \textbf{(A.1)}$$

with

$$A_1 = n^{-1/2}\sum_{i=1}^{n}\{f(X_i,\theta_0) - f(X_i,\theta_n)\}1(\beta' X_i \le u)$$

and

$$A_2 = n^{-1}\sum_{i=1}^{n} a(X_i)1(\beta' X_i \le u).$$

Using A3′ as in Theorem 2, we obtain

$$|A_1 + n^{1/2}(\theta_n - \theta_0)G(x,\theta_0)| = o_P(1)$$

https://doi.org/10.1017/S0266466606060506 Published online by Cambridge University Press

uniformly in $x \in \Pi$. On the other hand, using the results of Wolfowitz (1954), we have uniformly in $x \in \Pi$

$$|A_2 - E[a(X)1(\beta' X \leq u)]| = o_P(1).$$

Using the preceding equations and (A.1), the theorem holds from Theorem 1 and Assumption A3′. ■

APPENDIX B: Computation of the Test Statistic

By simple algebra

$$\begin{aligned} PCvM_n &= \int_{\Pi} |R_n^1(\beta,u)|^2 F_{n,\beta}(du)\, d\beta \\ &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} e_i(\theta_n) e_j(\theta_n) \int_{\Pi} 1(\beta' X_i \leq u) 1(\beta' X_j \leq u) F_{n,\beta}(du)\, d\beta \\ &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} e_i(\theta_n) e_j(\theta_n) \int_{\mathbb{S}_d} 1(\beta' X_i \leq \beta' X_r) 1(\beta' X_j \leq \beta' X_r)\, d\beta \\ &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} e_i(\theta_n) e_j(\theta_n) A_{ijr}. \end{aligned}$$

For $d > 1$, note that the integral A_{ijr} is proportional to the volume of a spherical wedge and hence we can compute them from the formula

$$A_{ijr} = A_{ijr}^{(0)} \frac{\pi^{(d/2)-1}}{\Gamma\left(\dfrac{d}{2}+1\right)},$$

where $A_{ijr}^{(0)}$ is the complementary angle between the vectors $(X_i - X_r)$ and $(X_j - X_r)$ measured in radians and $\Gamma(\cdot)$ is the gamma function. Thus, $A_{ijr}^{(0)}$ is given by

$$A_{ijr}^{(0)} = \left| \pi - ar\cos\left(\frac{(X_i - X_r)'(X_j - X_r)}{|(X_i - X_r)||(X_j - X_r)|} \right) \right|.$$

Hence, the computation of these integrals is simple. In addition, there are some restrictions on the integrals A_{ijr} that make the computation simpler, e.g., if $X_i = X_j$ and $X_i \neq X_r$ then $A_{ijr}^{(0)} = \pi$, whereas if $X_i = X_j$ and $X_i = X_r$ then $A_{ijr}^{(0)} = 2\pi$. If $X_i \neq X_j$ and $X_i = X_r$ or $X_j = X_r$, we have that $A_{ijr}^{(0)} = \pi$. Also, the symmetric property $A_{ijr} = A_{jir}$ holds.