ELSEVIER

Journal of Econometrics 98 (2000) 27–46

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Consistent bootstrap tests of parametric regression functions

Yoon-Jae Whang*

Department of Economics, Ewha University, Seoul 120-750, South Korea

Received 1 March 1998; received in revised form 12 July 1999; accepted 1 August 1999

Abstract

This paper introduces specification tests of parametric mean-regression models. The null hypothesis of interest is that the parametric regression function is correctly specified. The proposed tests are generalizations of the Kolmogorov–Smirnov and Cramer–von Mises tests to the regression framework. They are consistent against all alternatives to the null hypothesis, powerful against $1/\sqrt{n}$ local alternatives, not dependent on any smoothing parameters and simple to compute. A wild-bootstrap procedure is suggested to obtain critical values for the tests and is justified asymptotically. A small-scale Monte Carlo experiment shows that our tests (especially Cramer–von Mises test) have outstanding small sample performance compared to some of the existing tests. © 2000 Published by Elsevier Science S.A. All rights reserved.

JEL classification: C12

Keywords: Bootstrap; Consistent test; Parametric regression model; Specification test

1. Introduction

This paper proposes specification tests for a parametric mean-regression model for independent observations. The null hypothesis of interest (H_0) is that

* Tel.: + 82-2-3277-2793; fax: + 82-2-3277-2790 Web: http://mm.ewha.ac.kr/whang.
E-mail address: whang@mm.ewha.ac.kr (Y.-J. Whang)

0304-4076/00/$ - see front matter
PII: S0304-4076(99)00078-0

the parametric regression function is correctly specified. The alternative hypothesis (H_1) is the negation of the null hypothesis, i.e., that the parametric regression function is incorrectly specified. The parametric regression model we consider is a nonlinear regression model in which the regression function is known except for a finite-dimensional parameter and the regression errors may be heteroskedastic.

The tests we consider are generalizations of the Kolmogorov–Smirnov and Cramer–von Mises tests of goodness of fit. The tests introduced in this paper are (i) consistent against all alternatives to the null hypothesis H_0; (ii) powerful against $1/\sqrt{n}$ local alternatives to H_0; (iii) not dependent on any smoothing parameters; and (iv) simple to compute.

We note that there is a huge literature on the testing problem considered here. Some of the existing tests in the literature employ a nonparametric estimator of one sort or another of the regression function. Examples of such tests include Eubank and Spiegelman (1990), Fan and Li (1996a), Gozalo (1993), Härdle and Mammen (1993), Hong and White (1996), Li and Wang (1998), Whang and Andrews (1993), Wooldridge (1992), and Yatchew (1992), to mention only a few. These tests are consistent against all alternatives to the null, but are not powerful against $1/\sqrt{n}$ local alternatives and dependent on smoothing parameters. On the other hand, the tests of Bierens (1982, 1990), Bierens and Ploberger (1997), and De Jong (1996), among others, do not employ any nonparametric estimator of the regression function and hence do not suffer from the latter problems. These tests, however, have case-dependent asymptotic distributions and hence require a suitable choice of a nuisance parameter space, which could be arbitrary, and rely on an upper bound on the asymptotic critical value (in the cases of Bierens (1982) and Bierens and Ploberger (1997)), which might be too conservative. Our tests suggested below also have case-dependent asymptotic distributions, but try to avoid the above problems by using a bootstrap procedure to obtain critical values.

We also note that Andrews (1997, hereafter AN) has recently generalized the Kolmogorov–Smirnov test to develop a consistent specification test for parametric models that specify a parametric family for the conditional *distribution* of a response variable given a covariate. The present paper differs from AN in that it considers specification tests for parametric models that specify a parametric family for the regression function, that is, the conditional *mean* of the response variable given the covariate. Due to this difference, the parametric bootstrap procedure considered by AN is not appropriate for our purpose to simulate the asymptotic null distributions of the test statistics. Instead, it turns out that a wild-bootstrap procedure introduced in Section 4 below works in our context. Furthermore, this paper also considers a generalization of the Cramer–von Mises test to the regression context, in a hope that it might have better finite sample power performance than the Kolmogorov–Smirnov-type test. Except these differences, this paper follows the approach of AN very closely – we extensively use the notation of AN to help a reader familiar with the latter paper.

Finally, we note that specification tests similar to ours have been suggested independently by Stute et al. (1998a). Our paper, however, allows more general null models and estimation procedures than theirs and discusses power properties more explicitly.

The remainder of the paper is organized as follows. Section 2 defines the main test statistics. Section 3 establishes the asymptotic null distributions of the test statistics. Section 4 introduces a bootstrap procedure for obtaining critical values and justifies the procedure asymptotically. Section 5 establishes consistency of the tests. Section 6 determines the power of the tests against $1/\sqrt{n}$ local alternatives. An Appendix sketches proofs of results stated in the text.

2. Definition of the test statistics

Suppose that we observe a sample of n r.v.'s $\{(Y_i, X_i): i = 1, \ldots, n\}$, where $Y_i \in R$ and $X_i \in R^{K}$. We assume that the r.v.'s satisfy:[1]

Assumption D1. $\{(Y_i, X_i): i \geqslant 1\}$ are i.i.d. with conditional distribution function $H(\cdot|X_i)$ of Y_i given X_i and marginal distribution function $G(\cdot)$ of X_i.

The null hypothesis of interest is

$$\mathrm{H}_0:\quad P_G\left(\int y \, \mathrm{d}H(y|X) = g(X, \theta)\right) = 1 \quad \text{for some } \theta \in \Theta \subset R^P, \tag{1}$$

where $g(\cdot, \theta)$ is a known regression function and $X \sim G$. The alternative hypothesis is the negation of H_0, that is,

$$\mathrm{H}_1:\quad P_G\left(\int y \, \mathrm{d}H(y|X) \neq g(X, \theta)\right) > 0 \quad \text{for all } \theta \in \Theta \subset R^P. \tag{2}$$

We now define our test statistics. Let

$$\hat{H}_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i (X_i \leqslant x) \tag{3}$$

$$\hat{F}_n(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta)(X_i \leqslant x), \tag{4}$$

where $(X_i \leqslant x)$ denotes the indicator function of the event $X_i \leqslant x$. That is, $(X_i \leqslant x) = 1$ if $X_i \leqslant x$ and $(X_i \leqslant x) = 0$ otherwise. Let $\hat{\theta}$ be an estimator of θ.

[1] See Appendix for a discussion of results that are applicable to the case of independent non-identically distributed (i.n.i.d.) r.v.'s.

When H_0 is true, we let θ_0 denote the true value of θ. Note that under the null hypothesis $E(\hat{H}_n(x)|\mathbb{X}) = \hat{F}_n(x, \theta_0)\ \forall x \in R^K$, where $E(\cdot|\mathbb{X})$ denotes conditional expectation given $\mathbb{X} = \{X_i: i \geqslant 1\}$. Under the alternative hypothesis, however, $E(\hat{H}_n(x)|\mathbb{X}) \neq \hat{F}_n(x, \theta)$ for some $x \in R^K$, for all $\theta \in \Theta$ (see Section 5 below). Therefore, we take the difference between $\hat{H}_n(\cdot)$ and $\hat{F}_n(\cdot, \hat{\theta})$ as the basis of our test statistic.

The test statistics are defined by

$$KS_n = \sqrt{n} \sup_{x \in \mathscr{X}} |\hat{H}_n(x) - \hat{F}_n(x, \hat{\theta})| \tag{5}$$

$$CM_n = n \int (\hat{H}_n(x) - \hat{F}_n(x, \hat{\theta}))^2 \mathrm{d}G_n(x), \tag{6}$$

where $\mathscr{X} = \mathrm{supp}(G) \subset R^K$ denotes the support of the distribution function (df) G and $G_n(\cdot)$ denotes the empirical distribution of $\{X_i: i = 1, \ldots, n\}$. The tests KS_n and CM_n are generalizations of the Kolmogorov–Smirnov and Cramer–von Mises tests respectively to the regression context.[2]

The asymptotic distributions of KS_n and CM_n under the null hypothesis are turned out to be functionals of a Gaussian process that depends on nuisance parameters, viz. θ_0 and $G(\cdot)$ (see Section 3 below).[3] In consequence, we obtain critical values for the KS_n and CM_n statistics by a bootstrap procedure described in Section 4 below.

3. The asymptotic distributions of the tests under the null hypothesis

In this section, we examine the asymptotic distributions of the KS_n and CM_n statistics under the null hypothesis.

---

[2] Our tests as well as the Bierens-type tests can be interpreted as conditional moment tests, using either an indicator function or an exponential function as the weighting function. (See Stinchcombe and White (1998) for a general treatment of consistent conditional moment tests with a various choices of weighting functions.) As compared to the Bierens type tests, the choice of the indicator function as the weighting function has an advantage in that it does not require an arbitrary choice of a nuisance parameter space.

[3] Recently, Stute et al. (1998b) have also considered a test based on $\hat{H}_n(x) - F_n(x,\hat{\theta})$ in which X_i is assumed to be univariate. Contrary to ours, their test uses a martingale transformation that makes the asymptotic null distribution of their test statistic free of nuisance parameters. The martingale transformation, however, depends on the unknown variance of the regression errors and hence requires nonparametric estimation if the errors are (conditionally) heteroskedastic. Hence, contrary to ours, their test requires a choice of smoothing parameters, which may be arbitrary in practice. Furthermore, to overcome the difficulty that the martingale transformation becomes unstable at extreme values of x, one needs to choose a suitable compact interval to calculate their test statistic, which again could be arbitrary in practice (see Stute et al., 1998b, p. 1923).

To justify the bootstrap procedure discussed in Section 4, we show that the asymptotic results introduced in this section hold conditional on $\{X_i\colon i \geqslant 1\}$ with probability one. Such conditional results are stronger than the corresponding unconditional results.[4] For brevity, we let 'wp1' denote 'with probability one'.

Let

$$v_n(x, \theta) = \sqrt{n}(\hat{H}_n(x) - \hat{F}_n(x, \theta)). \tag{7}$$

Note that $v_n(\cdot, \theta)$ is a conditional empirical process since $\mathrm{E}(\hat{H}_n(x)|\mathbb{X}) = \hat{F}_n(x, \theta)$.

We assume that the estimator $\hat{\theta}$ of θ satisfies the following linear expansion:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_0 \psi(Y_i, X_i, \theta_0) + \mathrm{o_p}(1) \text{ conditional on } \mathbb{X} \text{ wp1}, \tag{8}$$

where D_0 is a nonrandom $P \times P$ matrix and $\psi(\cdot,\cdot,\cdot)\colon R^K \times R \times \Theta \to R^P$ is a function satisfying Assumption E1(ii) below. This assumption is quite general and is satisfied by a number of parametric estimators including maximum likelihood and nonlinear least squares estimators (see Remark 2 to Assumption E1 below).

We show below that the asymptotic null distributions of the test statistics KS_n and CM_n depend on that of $(v_n(\cdot, \theta_0), \sqrt{n}\bar{\psi}_n(\theta_0)')'$, which is a Gaussian process, where

$$\bar{\psi}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, X_i, \theta). \tag{9}$$

Let G^* be any distribution function on R^K. Let

$$C(x_1, x_2, \theta, G^*) = \iint \begin{pmatrix} (y - g(x, \theta))(x \leqslant x_1) \\ \psi(y, x, \theta) \end{pmatrix} \begin{pmatrix} (y - g(x, \theta))(x \leqslant x_2) \\ \psi(y, x, \theta) \end{pmatrix}' \mathrm{d}H(y|x)\,\mathrm{d}G^*(x). \tag{10}$$

Then, the covariance matrix of the asymptotic distribution of $(v_n(\cdot, \theta_0), \sqrt{n}\, \bar{\psi}_n(\theta_0)')'$ is defined to be

$$C(x_1, x_2, \theta_0, G) = \mathrm{Cov}((v_n(x_1, \theta_0), \sqrt{n}\, \bar{\psi}_n(\theta_0)')', (v_n(x_2, \theta_0), \sqrt{n}\, \bar{\psi}_n(\theta_0)')'). \tag{11}$$

[4] This follows from the bounded convergence theorem, see footnote 2 of AN for details.

Weak convergence of $(v_n(\cdot,\,\theta_0),\,\sqrt{n}\,\bar{\psi}_n(\theta_0)')'$ requires the specification of a pseudometric ρ on R^K. It is defined as follows:[5] For $x_1, x_2 \in R^K$,

$$\rho(x_1, x_2) = \left[\iint y^2[(x \leqslant x_1) - (x \leqslant x_2)]^2\,\mathrm{d}H(y|x)\,\mathrm{d}G(x)\right]^{1/2}. \tag{12}$$

We now introduce assumptions used to establish our asymptotic results.

Assumption M1. (i) $g(X_i,\,\theta)$ is differentiable in θ on a neighborhood N_1 of θ_0 $\forall i \geqslant 1$. (ii) $\sup_{x\in\mathscr{X}}\sup_{\theta:\|\theta-\theta_0\|\leqslant r_n}\|(1/n)\sum_{i=1}^n(\partial/\partial\theta)g(X_i,\,\theta)(X_i \leqslant x) - \Delta_0(x)\| \to 0$ wp1 for all sequences of positive constants $\{r_n\colon n \geqslant 1\}$ such that $r_n \to 0$, where $\Delta_0(x) = \int(\partial/\partial\theta)g(\tilde{x},\,\theta_0)(\tilde{x} \leqslant x)\mathrm{d}G(\tilde{x})$. (iii) $\sup_{x\in R^k}\|\Delta_0(x)\| < \infty$ and $\Delta_0(\cdot)$ is uniformly continuous on R^K (with respect to the metric ρ).

Assumption E1. (i) $\sqrt{n}(\hat{\theta} - \theta_0) = (1/\sqrt{n})\sum_{i=1}^n D_0\psi(Y_i,\,X_i,\,\theta_0) + o_p(1)$ conditional on $\mathbb{X}$ wp1, where D_0 is a nonrandom $P\times P$ matrix that may depend on θ_0. (ii) $\psi(y,\,x,\,\theta)$ is a measurable function from $R^K\times R\times\Theta$ to R^P that satisfies (a) $\psi(y,\,x,\,\theta) = (y - g(x,\,\theta))\xi(x,\,\theta)$ for some function $\xi(\cdot,\cdot)\colon R^K\times\Theta\to R^P$ and (b) $\int\|\psi(y, x,\,\theta_0)\|^{2+\varepsilon}\mathrm{d}H(y|X)\mathrm{d}G(x) < \infty$ for some $\varepsilon > 0$. (iii) $\int|y|^{2+\delta}\mathrm{d}H(y|x)\mathrm{d}G(x) < \infty$ for some $\delta > 0$.

Remark 1. Assumption M1 is trivially satisfied when the null model is the linear regression model provided a suitable moment condition holds for the regressor X_i.

Remark 2. Assumption E1 is also very general and can be verified for many parametric estimators that are $\sqrt{n}$-consistent and asymptotically normal using the results in the literature. For example, for the nonlinear least-squares (NLS) estimator, the function $\xi(x,\,\theta)$ corresponds to $\partial g(x,\,\theta)/\partial\theta$ and D_0 is the inverse of the outer product $\mathrm{E}\psi(Y_i,\,X_i,\,\theta_0)\psi(Y_i,\,X_i,\,\theta_0)'$.

We now establish the asymptotic null distribution of our test statistics. Let $(v(\cdot),\,v_0')'$ be a mean zero Gaussian process with covariance function defined by

$$\mathrm{E}\begin{pmatrix} v(x_1) \\ v_0 \end{pmatrix}\begin{pmatrix} v(x_2) \\ v_0 \end{pmatrix}' = C(x_1,\,x_2,\,\theta_0,\,G). \tag{13}$$

[5] This choice of pseudometric is to use the functional CLT result of Pollard (1990, Theorem 10.6) to prove Theorems A.1 and A.2 of Appendix, see Eq. (10.5) in p. 53 of the latter paper for the definition of the pseudometric.

The asymptotic null distributions of our test statistics are given by

Theorem 1. Suppose Assumptions D1, M1, *and* E1 *hold. Then, under the null,* (*a*) $KS_n \xrightarrow{d} \sup_{x \in \mathscr{X}} |v(x) - \Delta_0(x)' D_0 v_0|$ *conditional on* $\mathbb{X}$ *wp1 and* (*b*) $CM_n \xrightarrow{d} \int (v(x) - \Delta_0(x)' D_0 v_0)^2 \mathrm{d}G(x)$ *conditional on* $\mathbb{X}$ *wp1.*

Remark 1. The asymptotic null distributions of KS_n and CM_n depend on the 'true' parameter θ_0 and distribution function $G(\cdot)$. The latter implies that the asymptotic critical values for KS_n and CM_n cannot be tabulated.[6]

Remark 2. By the bounded convergence theorem, the unconditional asymptotic distributions of the test statistics under the null hypothesis are the same as the conditional asymptotic distributions given in Theorem 1.

4. Bootstrap critical values

The basic problem for bootstrapping a test statistic is how to impose the null hypothesis in the resampling scheme. That is, the essential problem is to find a bootstrap distribution that mimics the null distribution of the test statistic, even though the data fails to satisfy the null hypothesis. In this paper, we consider a wild bootstrap procedure that successfully imposes the null restriction and allows heteroskedastic errors (see Wu (1986), Härdle and Mammen (1993), and Li and Wang (1998) for more discussions about the wild bootstrap procedure).

The wild bootstrap procedure for the test KS_n (CM_n) is carried out as follows:

Step 1: Use the original sample $\{(Y_i, X_i): i \leqslant n\}$ to compute $\hat{\theta}$. Then, obtain $\hat{U}_i = Y_i - g(X_i, \hat{\theta})$ for $i = 1, \ldots, n$.

Step 2: Obtain the bootstrap error $\{U_i^* : i = 1, \ldots, n\}$ using a two point distribution, i.e., $U_i^* = a_1 \hat{U}_i$ with probability $r = (\sqrt{5} + 1)/(2\sqrt{5})$ and $U_i^* = a_2 \hat{U}_i$ with probability $1 - r$ for $i = 1, \ldots, n$, where $a_1 = -(\sqrt{5} - 1)/2$ and $a_2 = (\sqrt{5} + 1)/2$.

Step 3: Let $Y_i^* = g(X_i, \hat{\theta}) + U_i^*$ for $i = 1, \ldots, n$. The resulting sample $\{(Y_i^*, X_i): i \leqslant n\}$ is the bootstrap sample.

Step 4: Compute the bootstrap value of KS_n (CM_n) call it KS_n^* (CM_n^*), by applying the definition of KS_n (CM_n) to the bootstrap sample in place of the original sample.

[6] Park and Whang (1999) apply the idea of this paper to develop Kolmogorov–Smirov and Cramer–von Mises-type tests for the null of the martingale hypothesis. Contrary to this paper, their test statistics do not depend on any estimated parameter and the asymptotic null distributions are nuisance parameter-free and hence the critical values can be tabulated, see Figs. 1 and 2 of Park and Whang (1999) for the shape of the asymptotic null distributions.

Step 5: Repeat Steps 2–4 above B times to get a sample $\{KS_n^*\}$ ($\{CM_n^*\}$) of the bootstrap value of KS_n (CM_n). The distribution of this sample, call it the bootstrap distribution, mimics the null distribution of KS_n (CM_n).

Step 6: Let $c_{\alpha nB}^{KS}(\hat{\theta})$ ($c_{\alpha nB}^{CM}(\hat{\theta})$) be the $(1-\alpha)$th sample quantile of the bootstrap distribution of KS_n^* (CM_n^*). It is the *bootstrap critical value* of significance level α. Thus, we reject the null hypothesis at the significance level α if $KS_n > c_{\alpha nB}^{KS}(\hat{\theta})(CM_n > c_{\alpha nB}^{CM}(\hat{\theta}))$.

We now briefly discuss the asymptotic validity of the wild bootstrap procedure. (For a more complete discussion, see Whang, 1998). Let $H(\cdot,\cdot)$ denote the unconditional joint distribution function of (Y_i, X_i), i.e.,

$$H(y, x) = \int H(y|\tilde{x})(\tilde{x} \leqslant x)\mathrm{d}G(\tilde{x}) \quad \text{for } (y, x) \in R^{K+1}. \tag{14}$$

For any distribution function H^* on R^{K+1}, define

$$C^0(x_1, x_2, \theta, H^*) = \iint \begin{pmatrix} (y - g(x, \theta))(x \leqslant x_1) \\ \psi(y, x, \theta) \end{pmatrix} \begin{pmatrix} (y - g(x, \theta))(x \leqslant x_2) \\ \psi(y, x, \theta) \end{pmatrix}' \mathrm{d}H^*(y, x). \tag{15}$$

Note that $C^0(x_1, x_2, \theta, H) = C(x_1, x_2, \theta, G)$ under Assumption D1 (see Eq. (10) for the definition of $C(x_1, x_2, \theta, G)$).

For any sequence of nonrandom parameters $\{\theta_n : n \geqslant 1\}$ such that $\theta_n \to \theta_0$, let the boostrap sample $\{(Y_i^*, X_i) : i \leqslant n\}$ be distributed under $\{\theta_n : n \geqslant 1\}$ when the regressor $\{X_i : i \geqslant 1\}$ are i.i.d. with df $G(\cdot)$ and the regressand $\{Y_i^* : i \leqslant n, n \geqslant 1\}$ are independently generated from $Y_i^* = g(X_i, \theta_n) + U_i^*$ for $i = 1, \ldots, n$ in Step 3 above.

We assume that the estimator $\hat{\theta}_n^*$ of θ_n based on the bootstrap sample satisfies the following assumption under $\{\theta_n : n \geqslant 1\}$:

Assumption E2. (i) For all nonrandom sequence $\{\theta_n : n \geqslant 1\}$ for which $\theta_n \to \theta_0$, we have $\sqrt{n}(\hat{\theta}^* - \theta_n) = (1/\sqrt{n})\sum_{i=1}^n D_0 \psi(Y_i^*, X_i, \theta_n) + o_p(1)$ conditional on $(\mathbb{Y}, \mathbb{X}) = \{(Y_i, X_i) : i \geqslant 1\}$ wp1 for D_0 and $\psi(y, x, \theta)$ as defined in Assumption E1(ii). (ii) $\iint \tilde{\psi}_1(y, x)\,\mathrm{d}H(y|x)\,\mathrm{d}G(x) < \infty$, where $\tilde{\psi}_1(y, x) = \sup_{\theta \in N_1} \|\psi(y, x, \theta)\|^{2+\varepsilon}$ for some $\varepsilon > 0$. (iii) $\int |y|^{2+\delta}\,\mathrm{d}H(y|x)\,\mathrm{d}G(x) < \infty$, $\int \sup_{\theta \in N_1} |g(x, \theta)|^{2+\delta}\,\mathrm{d}G(x) < \infty$, and $\int \sup_{\theta \in N_1} \|(\partial/\partial\theta) g(x, \theta)\|^2\,\mathrm{d}G(x) < \infty$ for some $\delta > 0$.

In addition, we assume that the covariance matrix is continuous in θ

Assumption M2. $C^0(x_1, x_2, \theta, H)$ is continuous in θ at $\theta_0 \forall x_1, x_2 \in R^K$.

When the form of the regression function $\int y\,\mathrm{d}H(y|x)$ is arbitrary, we assume that the estimator $\hat{\theta}$ satisfies:

Assumption E3. $\hat{\theta} \to \theta_1$ a.s. conditional on $\mathbb{X}$ wp1 for some $\theta_1 \in \Theta$.

We now establish that the bootstrap distribution of KS_n^* (CM_n^*) has the same limit as the asymptotic null distribution of KS_n (CM_n) when the bootstrap sample is generated from $Y_i^* = g(X_i, \theta_n) + U_i^*$. That is, we have:

Theorem 2. Suppose Assumptions D1, M1, M2, and E2 hold. Then, for any nonrandom sequence $\{\theta_n: n \geqslant 1\}$ *for which* $\theta_n \to \theta_0$, *we have* (a) $KS_n^* \xrightarrow{\mathrm{d}} \sup_{x\in\mathcal{X}} |v(x) - \Delta_0(x)' D_0 v_0|$ *under* $\{\theta_n: n \geqslant 1\}$ *conditional on* $(\mathbb{Y}, \mathbb{X})$ *wp1 and* (b) $CM_n^* \xrightarrow{\mathrm{d}} \int (v(x) - \Delta_0(x)' D_0 v_0)^2\,\mathrm{d}G(x)$ *under* $\{\theta_n: n \geqslant 1\}$ *conditional on* $(\mathbb{Y}, \mathbb{X})$ *wp*1.

Given the absolute continuity of the limit distribution,[7] Theorem 2 suggests that the level α critical value $c_{\alpha n}^{KS}(\theta_n)$ ($c_{\alpha n}^{CM}(\theta_n)$) obtained from the bootstrap distribution of KS_n^* (CM_n^*) converges (conditional on the original sample) to the critical value $c_\alpha^{KS}(\theta_0)$ ($c_\alpha^{CM}(\theta_0)$) from the limit distribution of KS_n (CM_n) for any $\theta_0 \in \Theta$. This result, in turn, implies that, if $\hat{\theta}$ converges to some $\theta_1 (\in \Theta)$, then the critical value $c_{\alpha n}^{KS}(\hat{\theta})$ ($c_{\alpha n}^{CM}(\hat{\theta})$) from the bootstrap distribution with $Y_i^* = g(X_i, \hat{\theta}) + U_i^*$ converges (conditional on the original sample) to $c_\alpha^{KS}(\theta_1)$ ($c_\alpha^{CM}(\theta_1)$). The latter implies that asymptotic significance level of the test $KS_n(CM_n)$ with critical value $c_{\alpha n}^{KS}(\hat{\theta})$ ($c_{\alpha n}^{CM}(\hat{\theta})$) is α as desired.

The above heuristic arguments can be stated more formally in the following corollary (see Whang (1998) for a proof):

Corollary 3. (a) *Suppose Assumptions D1, M1, M2, E2, and E3* (*with* $\theta_1 = \theta_0$) *hold and* $B \to \infty$ *as* $n \to \infty$. *Then, under the null hypothesis,*

$$c_{\alpha n B}^{A}(\hat{\theta}) \to c_\alpha^{A}(\theta_0) \text{ a.s. conditional on } \mathbb{X} \text{ wp1 and}$$

$$\lim_{n\to\infty} P_{\theta_0}(A_n > c_{\alpha n B}^{A}(\hat{\theta})|\mathbb{X}) = \alpha \text{ wp1 for } A = KS \text{ and } CM.$$

(b) *Suppose Assumption D1 holds, Assumptions M1, M2, E2, and E3* (*with* $\theta_1 = \theta_0$) *hold for any value of* $\theta_0 \in \Theta$, *and* $B \to \infty$ *as* $n \to \infty$. *Then,*

$$\sup_{\theta_0 \in \Theta} \lim_{n\to\infty} P_{\theta_0}(A_n > c_{\alpha n B}^{A}(\hat{\theta})) = \alpha \text{ for } A = KS \text{ and } CM.$$

[7] The absolute continuity holds since the limit distribution is the supremum (or integral) of a Gaussian process whose covariance function is nonsingular, see Lifshits (1982).

(c) *Suppose Assumptions D1 and E3 hold, Assumptions M1, M2, and E2 hold for any value of* $\theta_0 \in \Theta$, *and* $B \to \infty$ *as* $n \to \infty$. *Then,*

$$c^{A}_{\alpha n B}(\hat{\theta}) \to c^{A}_{\alpha}(\theta_1) \text{ a.s. conditional on } \mathbb{X} \text{ wp1 for } A = KS \text{ and } CM.$$

Remark. Part (c) of Corollary 3 ensures that the tests KS_n and CM_n have power, since the bootstrap critical values converge almost surely to finite constants $c^{KS}_{\alpha}(\theta_1)$ and $c^{CM}_{\alpha}(\theta_1)$ respectively under Assumption E3 even when the null hypothesis fails to hold. (See Theorem 4 below.)

5. Consistency of the tests

In this section, we establish that the KS_n and CM_n tests are consistent against the alternative hypothesis H_1.

We assume that the following moment conditions hold under H_1:

Assumption E4. $\int\int y^2 \, dH(y|x) \, dG(x) < \infty$ and $\int |g(x, \theta_1)| \, dG(x) < \infty$.

Consistency of the tests is established in the following theorem:

Theorem 4. Suppose Assumptions D1, M1, E3 and E4 hold under the alternative hypothesis H_1. *Then, for all sequences of r.v.'s* $\{c_n\colon n \geqslant 1\}$ *with* $c_n = O_p(1)$ *conditional on* $\mathbb{X}$ *wp1, we have* $\lim_{n\to\infty} P(KS_n > c_n|\mathbb{X}) = 1$ *wp1 and* $\lim_{n\to\infty} P(CM_n > c_n|\mathbb{X}) = 1$ *wp1.*

Remark. The bootstrap critical value $c_{\alpha n B}(\hat{\theta})$ is $O_p(1)$ under the null and alternative by Corollary 3(c) and hence satisfies the requirements of Theorem 4 on c_n.

6. Local power properties

In this section, we determine the power of the KS_n and CM_n tests against local alternatives to the null hypothesis. The alternatives we consider are of distance $1/\sqrt{n}$ from the null hypothesis.

Suppose one is interested in the power of the KS_n and CM_n tests for the sample size n_0 against an alternative regression function $q(\cdot) \notin \{g(\cdot, \theta)\colon \theta \in \Theta\}$. Let $Q(\cdot|\cdot)$ denote a conditional distribution function of Y_i given X_i corresponding to $q(\cdot)$, i.e., $\int y \, dQ(y|x) = q(x)$. We create local alternatives by starting with the alternative regression function $q(\cdot)$ at the sample size n_0 and shrinking it toward $g(x, \theta_0)$ as the sample size increases. For this purpose, let

$$d(x) = \sqrt{n_0}(q(x) - g(x, \theta_0)). \tag{16}$$

Then the sequence of local alternative regression functions we consider is given by

$$q_n(x) = g(x, \theta_0) + d(x)/\sqrt{n} \quad \text{for } n = 1, 2, \ldots. \tag{17}$$

We denote $Q_n(\cdot|\cdot)$ to be a conditional distribution function corresponding to $q_n(\cdot)$, i.e., $\int y \, \mathrm{d}Q_n(y|x) = q_n(x)$. Below, we say the observations are distributed under $\{Q_n(\cdot|\cdot): n \geqslant 1\}$ when $\{(Y_i, X_i): i \leqslant n\}$ are distributed independently with conditional mean of Y_i given X_i given by $\int y \, \mathrm{d}Q_n(y|X_i) = q_n(X_i)$ for $i \leqslant n$ and $n \geqslant 1$. Note that the n_0th regression function is given by $\int y \, \mathrm{d}Q_{n_0}(y|X_i) = \int y \, \mathrm{d}Q(y|X_i) = q(X_i)$ as desired.

For our asymptotic results given below, we assume that the following moment conditions hold:

Assumption L. (i) $\int |q(x) - g(x, \theta_0)|^{2+\delta} \, \mathrm{d}G(x) < \infty$ for some $\delta > 0$.
(ii) $\int \|\psi(y, x, \theta_0)\|^2 \, \mathrm{d}Q(y|x) \, \mathrm{d}G(x) < \infty$.

We also assume that the 'regression errors' under the sequence of local alternatives have the same conditional moment (up to a constant) as that under the null hypothesis:

Assumption V. $\int (y - \int y \, \mathrm{d}Q_n(y|x))^r \, \mathrm{d}Q_n(y|x) = \int (y - \int y \, \mathrm{d}H(y|x))^r \, \mathrm{d}H(y|x) \; \forall 0 < r \leqslant 2 + \varepsilon$ for some $\varepsilon > 0 \; \forall x \in R^K \; \forall n \geqslant 1$.

Remark. Assumption V is weaker than the corresponding assumption of Bierens and Ploberger (1997, p. 1132) who assume that the regression errors (themselves, not conditional moments of them) under local alternatives are the same as those under the null hypothesis.

Define

$$M_1 = \sup_{x \in \mathscr{X}} |v(x) - \Delta_0(x)' D_0 v_0 + \delta(x)| \tag{18}$$

and

$$M_2 = \int (v(x) - \Delta_0(x)' D_0 v_0 + \delta(x))^2 \, \mathrm{d}G(x), \tag{19}$$

where

$$\delta(x) = \sqrt{n_0} \left[\int (q(\tilde{x}) - g(\tilde{x}, \theta_0))(\tilde{x} \leqslant x) \, \mathrm{d}G(\tilde{x}) - \Delta_0(x)' D_0 \iint \psi(\tilde{y}, \tilde{x}, \theta_0) \, \mathrm{d}Q(\tilde{y}|\tilde{x}) \, \mathrm{d}G(\tilde{x}) \right].$$

The asymptotic distributions of KS_n and CM_n under the local alternatives are given by

Theorem 5. Suppose Assumptions M1, E1, L, and V hold. Then, (a) $KS_n \xrightarrow{d} M_1$ *under* $\{Q_n(\cdot|\cdot): n \geqslant 1\}$ *conditional on* $\mathbb{X}$ *wp*1 *and* (b) $CM_n \xrightarrow{d} M_2$ *under* $\{Q_n(\cdot|\cdot): n \geqslant 1\}$ *conditional on* $\mathbb{X}$ *wp*1.

Remark 1. By the same arguments to those in Andrews (1997, Comment 2 to Theorem 4, p. 1114), the KS_n and CM_n tests are asymptotically locally unbiased.

Remark 2. In contrast to the existing consistent tests that depend on a nonparametric estimator of the regression function (e.g., Härdle and Mammen, 1993; Fan and Li, 1996a; Hong and White, 1996), the KS_n and CM_n tests have nontrivial power against the $1/\sqrt{n}$ local alternatives of the form (17). It should be noted, however, that this result does not generally extend to hold uniformly over all sequences of local alternatives that are of distance $1/\sqrt{n}$ from the null model.[8]

Remark 3. The asymptotic local power of the KS_n and CM_n tests against $\{q_n(\cdot): n \geqslant 1\}$ is given by $P(M_1 > c_\alpha(\theta_0))$ and $P(M_2 > c_\alpha(\theta_0))$ respectively. Therefore, since $q_{n_0}(\cdot) = q(\cdot)$, the powers of the KS_n and CM_n tests against $q(\cdot)$ when the sample size is n_0 can be approximated by $P(M_1 > c_\alpha(\theta_0))$ and $P(M_2 > c_\alpha(\theta_0))$ respectively.

7. Monte Carlo experiment

In this section, we examine the finite sample performance of the tests KS_n and CM_n in a small scale simulation experiment. We compare the performance of our tests to those of Härdle and Mammen (1993) (hereafter HM) and Bierens and Ploberger (1997) (hereafter BP).

We considered two data generating processes: For $i = 1, \ldots, n$,

$$DGP1: Y_i = 2X_{i1} - X_{i1}^2 + c_1(X_{i1} - \tfrac{1}{4})(X_{i1} - \tfrac{1}{2})(X_{i1} - \tfrac{3}{4}) + U_{i1} \tag{20}$$

[8] Fan and Li (1996b), for example, show that the smoothing based tests can be more powerful than the nonsmoothing based tests for the 'singular' alternatives that converge to the null model at a rate faster than $1/\sqrt{n}$. Horowitz and Spokoiny (1999) further develop a test that is uniformly consistent against alternatives that converge to the null model at the fastest possible rate which is slower than $1/\sqrt{n}$.

Table 1
Rejection probabilities of the tests at the 5% significance levels (DGP1)[a]

c_1	Test	n		
		50	100	200
	KS	0.055	0.047	0.051
	CM	0.047	0.054	0.052
	$HM1$	0.010	0.019	0.026
	$HM2$	0.026	0.028	0.036
0.0	$HM3$	0.036	0.039	0.045
	$BP1$	0.025	0.033	0.031
	$BP2$	0.030	0.039	0.031
	KS	0.802 [0.880]	0.983 [0.994]	1.00 [1.000]
	CM	0.884 [0.927]	0.998 [1.000]	1.00 [1.000]
	$HM1$	0.199 [0.828]	0.355 [0.989]	0.517 [1.000]
0.05	$HM2$	0.445 [0.877]	0.598 [0.996]	0.679 [1.000]
	$HM3$	0.604 [0.895]	0.702 [0.999]	0.778 [1.000]
	$BP1$	0.882 [0.937]	0.962 [0.987]	0.973 [0.985]
	$BP2$	0.904 [0.939]	0.982 [0.997]	0.995 [1.000]
	KS	0.986 [0.995]	0.999 [1.000]	1.00 [1.000]
	CM	0.992 [0.998]	1.00 [1.000]	1.00 [1.000]
	$HM1$	0.383 [0.989]	0.446 [1.000]	0.544 [1.000]
0.1	$HM2$	0.668 [0.997]	0.657 [1.000]	0.690 [1.000]
	$HM3$	0.790 [0.997]	0.748 [1.000]	0.791 [1.000]
	$BP1$	0.973 [0.992]	0.974 [0.992]	0.973 [0.984]
	$BP2$	0.980 [0.991]	0.991 [0.999]	0.995 [1.000]

[a] $HM1$, $HM2$, and $HM3$ represent the Härdle and Mammen's (1993) test with bandwidth parameter $h = 0.1$, 0.2, and 0.3, respectively. $BP1$ and $BP2$ correspond to the Bierens and Ploberger's (1990) test with the nuisance parameter space given by $\Xi = [1, 3]$ and $[1, 5]$, respectively. The numbers in the square brackets are rejection probabilities using the empirical critical values.

and

$$DGP2: Y_i = 1 + X_{i1} + X_{i2} + c_2 X_{i1} X_{i2} + U_{i2},$$

where X_{i1} and $X_{i2} \sim$ i.i.d. N(0, 1), $U_{i1} \sim$ i.i.d. N(0, 0.01) and $U_{i2} \sim$ i.i.d. N(0, 4). We take $c_1 \in (0.0, 0.05, 0.1)$ and $c_2 \in (0.0, 1.0, 2.0)$. The null models correspond to $c_1 = 0.0$ and $c_2 = 0.0$. DGP1 is the same as the model used by HM (except that X_{i1} are generated from the uniform distribution in the latter) and DGP2 is similar to the design considered by Bierens (1990).

We consider three sample sizes: $n \in (50, 100, 200)$. The number of bootstrap repetitions is $B = 200$. The number of Monte Carlo simulation repetitions is

Table 2
Rejection probabilities of the tests at the 5% significance levels (DGP2)[a]

c_2	Test	n		
		50	100	200
	KS	0.061	0.063	0.061
	CM	0.048	0.057	0.057
	$HM1$	0.010	0.008	0.023
0.0	$HM2$	0.015	0.010	0.023
	$HM3$	0.031	0.037	0.064
	$BP1$	0.018	0.016	0.027
	$BP2$	0.010	0.009	0.016
	KS	0.232 [0.374]	0.494 [0.656]	0.835 [0.919]
	CM	0.393 [0.559]	0.762 [0.831]	0.989 [0.995]
	$HM1$	0.035 [0.506]	0.173 [0.818]	0.557 [0.985]
1.0	$HM2$	0.062 [0.537]	0.233 [0.834]	0.660 [0.989]
	$HM3$	0.105 [0.543]	0.338 [0.851]	0.729 [0.992]
	$BP1$	0.235 [0.381]	0.570 [0.734]	0.900 [0.953]
	$BP2$	0.108 [0.290]	0.306 [0.585]	0.622 [0.805]
	KS	0.501 [0.887]	0.845 [0.991]	0.994 [1.000]
	CM	0.811 [0.961]	0.991 [1.000]	1.00 [1.000]
	$HM1$	0.139 [0.949]	0.449 [1.000]	0.873 [1.000]
2.0	$HM2$	0.215 [0.952]	0.539 [1.000]	0.901 [1.000]
	$HM3$	0.257 [0.953]	0.570 [1.000]	0.880 [1.000]
	$BP1$	0.546 [0.715]	0.826 [0.912]	0.967 [0.984]
	$BP2$	0.300 [0.586]	0.560 [0.778]	0.756 [0.876]

[a] $HM1$, $HM2$, and $HM3$ represent the Härdle and Mammen's (1993) test with bandwidth parameter $h = 0.7$, 0.8, and 0.9, respectively. $BP1$ and $BP2$ correspond to the Bierens and Ploberger's (1990) test with the nuisance parameter space given by $\Xi = [1, 3] \times [1, 3]$ and $[1, 5] \times [1, 5]$, respectively. The numbers in the square brackets are rejection probabilities using the empirical critical values.

fixed to be 1000. Computational requirements prevented us from considering larger sample sizes and number of repetitions.

To calculate the HM test statistic, one needs to choose the kernel function $K(\cdot)$ and the bandwidth parameter (h). We use the quartic kernel $K(u) = (15/16)(1 - u^2)^2$ for $|u| \leqslant 1$ as in HM and we take $h \in (0.1, 0.2, 0.3)$ for DGP1 and $h \in (0.7, 0.8, 0.9)$ for DGP2. The critical values for the HM test are obtained from the wild bootstrap procedure as suggested by HM to improve the finite sample performance of their test. For BP test, on the other hand, one need to choose the nuisance parameter space Ξ (using the notation of BP). We take $\Xi \in ([1, 3], [1, 5])$ for DGP1 and $\Xi \in ([1, 3] \times [1, 3], [1, 5] \times [1,5])$ for DGP2. We use a numerical integration to calculate the BP test statistic.

Tables 1 and 2 provide the rejection probabilities of the four alternative tests with size α. The significance level considered was $\alpha = 0.05$. (Results at other significance levels were similar and hence are not reported here.) The numbers given in square brackets are size-corrected powers, i.e., rejection probabilities using empirical critical values. The empirical critical values are obtained from 1000 random samples of size n from the null models (i.e., DGP1 and DGP2 with $c_1 = 0$ and $c_2 = 0$ respectively).[9]

Table 1 shows the simulation results for DGP1. When $c_1 = 0$, the results indicate that the size performance of the KS test is reasonably good, whereas the HM and BP tests generally under-reject the null hypothesis. The under-rejection of the BP test is as expected since it relies on the upper bound on the asymptotic critical value. Also as expected, the performance of the HM test is quite sensitive to the choice of the bandwidth parameter. When $c_1 \neq 0$, the results show that the rejection probabilities increase as either c_1 or n increases. In terms of (both actual and size-corrected) power, the CM test has good performance compared to the other alternative tests in most of the cases considered. For the small sample $n = 50$ and $\Xi = [1, 5]$, however, the BP test has relatively good power performance.

Table 2 provides the corresponding results to DGP2, in which the dimension of the regressor is $K = 2$. The results show that there are substantial small sample size distortions for the test HM, which might be possibly due to the well-known curse of dimensionality problem of nonparametric estimators. The BP test now under-rejects the null hypothesis more severely than under DGP1. When $c_2 \neq 0$, the results clearly indicate that the CM test dominates the other competing tests.

To summarize, we find that the CM test has generally the best performance in our simulation experiments and we recommend practitioners to use the CM test in their future research.

Acknowledgements

I thank the associate editor, three anonymous referees, Don Andrews, M.-J. Lee, and the seminar participants of The 1999 Far Eastern Meeting of the Econometric Society held at Singapore for helpful comments.

This work was supported by the Brain Korea 21 Project.

[9] In practice, the empirical critical values are not available and hence might not be useful in applications. They are useful, however, in comparing power performance of the competing tests that might have size distortions in finite samples.

Appendix

The asymptotic results in the main text can be proved based on the ideas of AN. The main technical difficulty we have is to prove how the wild bootrap (rather than parametric bootstrap as considered by AN) works in our case. (Following the suggestion of a referee, we cut out the detailed treatment, which can be obtained from an earlier version of this paper, i.e. Whang (1998).

The main step is as follows: First, show that under suitable assumptions the desired asymptotic results of Theorem 1, 4, and 5 (Theorem 2) hold for a fixed (nonrandom) sequences $\{X_i : i \geqslant 1\}(\{(Y_i, X_i) : i \geqslant 1\})$. Next, verify that these assumptions hold with probability one if the corresponding assumptions in the Theorems 1, 2, 4, and 5 hold. This establishes the desired results in the text. Below, we omit the second step (which can be established by extending the results of Pollard (1984, Theorem II.24) and applying the Kolmogorov's Strong Law of Large Numbers) and merely sketch the first step.[10]

We first provide assumptions under which KS_n and CM_n have the desired asymptotic null distributions when the regressors are fixed:

Assumption F.D1. $\{X_i : i \geqslant 1\}$ are fixed (i.e., nonrandom) and $\{Y_i : i \geqslant 1\}$ are independent with conditional distribution function $H(\cdot|X_i)$ of Y_i given X_i for all $i \geqslant 1$.

Assumption F.C1. (i) $\hat{G}_n(x) \to G(x)\ \forall x \in R^K$ for some distribution function $G(\cdot)$. (ii) $\mathrm{supp}(\hat{G}_n) \subset \mathrm{supp}(G)\ \forall n \geqslant 1$. (iii) $\sup_{x \in \mathscr{X}} |\int\int \tilde{y}^2(\tilde{x} \leqslant x)\, \mathrm{d}H(\tilde{y}|\tilde{x})(\mathrm{d}\hat{G}_n(\tilde{x}) - \mathrm{d}G(\tilde{x}))| \to 0$. (iv) $C(x_1, x_2, \theta_0, \hat{G}_n) \to C(x_1, x_2, \theta_0, G)\ \forall x_1, x_2 \in R^K$.

Assumption F.M1. (i) $g(X_i, \theta)$ is differentiable in θ on a neighborhood N_1 of $\theta_0\ \forall i \geqslant 1$. (ii) $\sup_{x \in \mathscr{X}} \sup_{\theta:\ \|\theta - \theta_0\| \leqslant r_n} \|(1/n)\sum_{i=1}^n (\partial/\partial\theta) g(X_i, \theta)(X_i \leqslant x) - \Delta_0(x)\| \to 0$ for all sequences of positive constants $\{r_n : n \geqslant 1\}$ such that $r_n \to 0$, where $\Delta_0(x) = \int (\partial/\partial\theta) g(\tilde{x}, \theta)(\tilde{x} \leqslant x)\, \mathrm{d}G(\tilde{x})$. (iii) $\sup_{x \in R^k} \|\Delta_0(x)\| < \infty$ and $\Delta_0(\cdot)$ is uniformly continuous on R^K (with respect to the metric ρ).

Assumption F.E1. (i) $\sqrt{n}(\hat{\theta} - \theta_0) = (1/\sqrt{n})\sum_{i=1}^n D_0 \psi(Y_i, X_i, \theta_0) + o_p(1)$, where D_0 is a nonrandom $P \times P$ matrix that may depend on θ_0. (ii) $\psi(y, x, \theta)$ is a measurable function from $R^{K+1} \times \Theta$ to R^P that satisfies (a)

[10] The results of the first step can be used to establish the asymptotic validity of the KS_n and CM_n tests when the observations are i.n.i.d. In this case, the limit distribution functions $G(\cdot)$ and $H(\cdot,\cdot)$ equals $\lim_{n\to\infty}(1/n)\sum_{i=1}^n G_i(\cdot)$ and $\lim_{n\to\infty}(1/n)\sum_{i=1}^n H_i(\cdot,\cdot)$ respectively, where $X_i \sim G_i(\cdot)$ and $(Y_i, X_i) \sim H_i(\cdot,\cdot)\ \forall i \geqslant 1$.

$\psi(y, x, \theta) = (y - g(x, \theta))\xi(x, \theta)$ for some function $\xi(\cdot,\cdot)$: $R^K \times \Theta \to R^P$ and (b) $\overline{\lim}_{n\to\infty}\ (1/n)\sum_{i=1}^{n}\tilde{\psi}_0(X_i) < \infty$, where $\tilde{\psi}_0(x) = \int\|\psi(y, x, \theta_0)\|^{2+\varepsilon}\mathrm{d}H(y|x)$ for some $\varepsilon > 0$. (iii) $\overline{\lim}_{n\to\infty}\ (1/n\sum_{i=1}^{n}\zeta_0(X_i) < \infty$, where $\zeta_0(x) = \int|y|^{2+\delta}\mathrm{d}H(y|x)$ for some $\delta > 0$.

Theorem A.1. Suppose Assumptions F.D1, F.C1, F.M1, and E1 hold. Then, under the null, (a) $KS_n \xrightarrow{d} \sup_{x\in\mathscr{X}}|v(x) - \Delta_0(x)'D_0 v_0|$ *and* (b) $CM_n \xrightarrow{d} \int(v(x) - \Delta_0(x)'D_0 v_0)^2\mathrm{d}G(x)$.

Proof of Theorem A.1. The proof of Theorem A.1 is similar to that of Theorem A.2 below with Y_i^*, $\hat{\theta}^*$, and θ_n replaced by Y_i, $\hat{\theta}$, and θ_0, respectively, and the results conditioned on $\mathbb{X}$ instead of $(\mathbb{Y}, \mathbb{X})$. □

Next, we establish the bootstrap results when the original sample is nonrandom:

Assumption F.D2. $\{(Y_i, X_i)$: $i \geqslant 1\}$ are fixed, i.e., nonrandom.

Assumption F.C2. (i) $\hat{H}_n(y,x) \to H(y,x)\ \forall (y,x) \in R \times R^K$ for some distribution function $H(\cdot,\cdot)$. (ii) $\mathrm{supp}(\hat{H}_n) \subset \mathrm{supp}(H)\ \forall n \geqslant 1$. (iii) $\sup_{x\in\mathscr{X}}|\int\int y^2(\tilde{x} \leqslant x)(\mathrm{d}\hat{H}_n(y, \tilde{x}) - \mathrm{d}H(y, \tilde{x}))| \to 0$. (iv) $\sup_{x\in\mathscr{X}}|\int\int g(\tilde{x}, \theta_0)(y - g(\tilde{x}, \theta_0))(\tilde{x} \leqslant x)(\mathrm{d}\hat{H}_n(y, \tilde{x}) - \mathrm{d}H(y, \tilde{x}))| \to 0$. (v) For all nonrandom sequence $\{\theta_n : n \geqslant 1\}$ for which $\theta_n \to \theta_0$, $C^0(x_1, x_2, \theta_n, \hat{H}_n) - C^0(x_1, x_2, \theta_n, H) \to 0\ \forall x_1, x_2 \in R^K$.

Assumption F.E2. (i) For all nonrandom sequence $\{\theta_n$: $n \geqslant 1\}$ for which $\theta_n \to \theta_0$, we have $\sqrt{n}(\hat{\theta}^* - \theta_n) = (1/\sqrt{n})\sum_{i=1}^{n} D_0\psi(Y_i^*, X_i, \theta_n) + \mathrm{o_p}(1)$ for D_0 and $\psi(y, x, \theta)$ as defined in Assumption F.A3. (ii) $\overline{\lim}_{n\to\infty}\ (1/n)\sum_{i=1}^{n}\tilde{\psi}_1(Y_i, X_i) < \infty$. (iii) $\overline{\lim}_{n\to\infty}(1/n)\sum_{i=1}^{n}|Y_i|^{2+\delta} < \infty$, $\overline{\lim}_{n\to\infty}\ (1/n)\ \sum_{i=1}^{n}\sup_{\theta\in N_1}|g(X_i, \theta)|^{2+\delta} < \infty$ and $\overline{\lim}_{n\to\infty}\ (1/n\ )\sum_{i=1}^{n}\sup_{\theta\in N_1}\|(\partial/\partial\theta)g(X_i, \theta)\|^2 < \infty$ for some $\delta > 0$.

Theorem A.2. Suppose Assumptions F.D2, F.C2, F.M1, M2, and F.E2 hold. Then, for any nonrandom sequence $\{\theta_n$: $n \geqslant 1\}$ *for which* $\theta_n \to \theta_0$, *we have* (a) $KS_n^* \xrightarrow{d} \sup_{x\in\mathscr{X}}|v(x) - \Delta_0(x)'D_0 v_0|$ *under* $\{\theta_n : n \geqslant 1\}$ *and* (b) $CM_n^* \xrightarrow{d} \int(v(x) - \Delta_0(x)'D_0 v_0)^2\,\mathrm{d}G(x)$ *under* $\{\theta_n : n \geqslant 1\}$.

Proof of Theorem A.2. Let

$$\bar{\psi}_n^*(\theta) = \frac{1}{n}\sum_{i=1}^{n}\psi(Y_i^*, X_i, \theta), \tag{A.1}$$

$$v_n^*(x, \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[Y_i^* - g(X_i, \hat{\theta})](X_i \leqslant x). \tag{A.2}$$

Using a mean value expansion argument, we have

$$\sqrt{n}\sup_{x\in\mathscr{X}}|\hat{F}_n(x,\hat{\theta}^*)-\hat{F}_n(x,\theta_n)-\Delta_0(x)'D_0\bar{\psi}_n^*(\theta_n)| = \mathrm{o_p}(1) \text{ under } \{\theta_n : n\geqslant 1\}. \tag{A.3}$$

under Assumptions F.M1, F.D2, and F.E2 for any nonrandom sequence $\{\theta_n \colon n\geqslant 1\}$ for which $\theta_n\to\theta_0$. By the functional CLT of Pollard (1990, Theorems 10.2 and 10.6), we also have

$$\begin{pmatrix} v_n^*(\cdot,\theta_n) \\ \sqrt{n}\;\bar{\psi}_n^*(\theta_n) \end{pmatrix} \Rightarrow \begin{pmatrix} v(\cdot) \\ v_0 \end{pmatrix}, \tag{A.4}$$

if Assumptions F.D2, F.C2, F.E2, and M2 hold under $\{\theta_n : n\geqslant 1\}$, where $(v_n^*(\cdot,\theta_n),\sqrt{n}\;\bar{\psi}_n^*(\theta_n)')'$ is a stochastic process indexed by $x\in\mathscr{X}$ and the pseudometric ρ on $\mathscr{X}$ is given by

$$\rho(x_1,x_2)=\left[\iint y^2[(x\leqslant x_1)-(x\leqslant x_2)]^2\,\mathrm{d}H(y,x)\right]^{1/2}. \tag{A.5}$$

Theorem A.2 now follows by combining the results of (A.3) and (A.4) and applying the continuous mapping theorem (see Pollard (1984, Theorem IV.12, p. 70)).

To obtain consistency of KS_n in the fixed regressors case, we assume

Assumption F.C3. (i) $\sup_{x\in\mathscr{X}}|\iint y(\tilde{x}\leqslant x)\mathrm{d}H(y|\tilde{x})(\mathrm{d}\hat{G}_n(\tilde{x})-\mathrm{d}G(\tilde{x}))|\to 0$.
(ii) $\sup_{x\in\mathscr{X}}|\int g(\tilde{x},\theta_1)(\tilde{x}\leqslant x)(\mathrm{d}\hat{G}_n(\tilde{x})-\mathrm{d}G(\tilde{x}))|\to 0$.

The behavior of $\hat{\theta}$ under arbitrary regression function $\int y\,\mathrm{d}H(y|\cdot)$ is given by

Assumption F.E4. (i) $\hat{\theta}\to\theta_1$ a.s. for some $\theta_1\in\Theta$. (ii) $\iint y^2\,\mathrm{d}H(y|x)\,\mathrm{d}G(x) <\infty\ \ \forall i\geqslant 1$.

Theorem A.3. Suppose Assumptions F.D1, F.C3, F.M1 and *F.E4 hold under the alternative hypothesis* $H1$. *Then, for all sequences of r.v.'s* $\{c_n \colon n\geqslant 1\}$ *with* $c_n=\mathrm{O_p}(1)$, we have $\lim_{n\to\infty}P(KS_n>c_n)=1$ *and* $\lim_{n\to\infty}P(CM_n>c_n)=1$.

Proof of Theorem A.3. Let

$$H(x)=\iint y(\tilde{x}\leqslant x)\,\mathrm{d}H(y|\tilde{x})\,\mathrm{d}G(\tilde{x}), \tag{A.6}$$

$$F(x,\theta)=\int g(\tilde{x},\theta)(\tilde{x}\leqslant x)\,\mathrm{d}G(\tilde{x}). \tag{A.7}$$

By Assumptions F.C3, F.E4, and the empirical process uniform LLN of Pollard (1990, Theorem 8.3), we have

$$\frac{1}{\sqrt{n}} KS_n = \sup_{x \in \mathscr{X}} |H(x) - F(x, \theta_1)| + o_p(1),$$

$$\frac{1}{n} CM_n = \int (H(x) - F(x, \theta_1))^2 \, dG(x) + o_p(1).$$

Theorem A.4 follows since there exists a point $x \in \mathscr{X}$ for which

$$H(x) - F(x, \theta_1) = \int \left[\int y \, dH(y|\tilde{x}) - g(\tilde{x}, \theta_1) \right] (\tilde{x} \leqslant x) \, dG(\tilde{x}) \neq 0. \tag{A.8}$$

under the alternative hypothesis H_1. □

Finally, we establish the local power results when the regressors are fixed:

Assumption F.L. (i) $\overline{\lim}_{n \to \infty} (1/n) \sum_{i=1}^n |q(X_i) - g(X_i, \theta_0)|^{2+\delta} < \infty$ for some $\delta > 0$.
(ii) $\overline{\lim}_{n \to \infty} (1/n) \sum_{i=1}^n \int \|\psi(y, X_i, \theta_0)\|^2 \, dQ(y|X_i) < \infty$.
(iii) $\sup_{x \in \mathscr{X}} |\int (q(\tilde{x}) - g(\tilde{x}, \theta_0))(\tilde{x} \leqslant x)(d\hat{G}_n(\tilde{x}) - dG(\tilde{x}))| \to 0$.
(iv) $\sup_{x \in \mathscr{X}} |\int (g(\tilde{x}, \theta_0))^2 (\tilde{x} \leqslant x)(d\hat{G}_n(\tilde{x}) - dG(\tilde{x}))| \to 0$.
(v) $\int\int \psi(y, x, \theta_0)(\tilde{x} \leqslant x) \, dQ(y|x)(d\hat{G}_n(\tilde{x}) - dG(\tilde{x})) \to 0$.

Theorem A.4. Suppose Assumptions F.C1, F.M1, F.E1, F.L, and V hold. Then, (a) $KS_n \xrightarrow{d} M_1$ *under* $\{Q_n(\cdot|\cdot): n \geqslant 1\}$ *and* (b) $CM_n \xrightarrow{d} M_2$ *under* $\{Q_n(\cdot|\cdot): n \geqslant 1\}$.

Proof. Under $\{Q_n(\cdot|\cdot): n \geqslant 1\}$, we have

$$\sqrt{n} \sup_{x \in \mathscr{X}} |\hat{F}_n(x, \hat{\theta}) - \hat{F}_n(x, \theta_0) - \Delta_0(x)' D_0 \bar{\psi}_n(\theta_0)| = o_p(1). \tag{A.9}$$

We also have

$$\begin{pmatrix} v_n(\cdot, \theta_0) \\ \sqrt{n} \, \bar{\psi}_n(\theta_0) \end{pmatrix} \Rightarrow \begin{pmatrix} v(\cdot) + \sqrt{n_0} \int (q(\tilde{x}) - g(\tilde{x}, \theta_0))(\tilde{x} \leqslant \cdot) \, dG(\tilde{x}) \\ v_0 + \sqrt{n_0} \int\int \psi(\tilde{y}, \tilde{x}, \theta_0) \, dQ(\tilde{y}|\tilde{x}) dG(\tilde{x}) \end{pmatrix} \tag{A.10}$$

under $\{Q_n(\cdot|\cdot): n \geqslant 1\}$. Combining the above results and using the continuous mapping theorem give Theorem A.4. □

References

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.
Bierens, H.J., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.

Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1151.

De Jong, R.M., 1996. The Bierens test under data dependence. Journal of Econometrics 72, 1–32.

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

Fan, Y., Li, Q., 1996a. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Li, Q., 1996b. Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. Unpublished manuscript. Department of Economics, Uinversity of Windsor.

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

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Hong, Y., White, H., 1996. Consistent specification testing via nonparametric series regressions. Econometrica 63, 1133–1160.

Horowitz, J.L., Spokoiny, V.G., 1999. An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. Unpublished manuscript. Department of Economics, University of Iowa.

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

Lifshits, M.A., 1982. On the absolute continuity of distributions of functionals of random processes. Theory of Probability and Its Applications 27, 600–607.

Park, J., Whang, Y.-J., 1999. Testing for the martingale hypothesis. Unpublished manuscript.

Pollard, D., 1984. Convergence of stochastic processes. Springer, New York.

Pollard, D., 1990. Empirical processes: theory and applications. CBMS Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward.

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

Stute, W., Manteiga, W.G., Quindmill, M.P., 1998a. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141–149.

Stute, W., Thies, S., Zhu, L.-X., 1998b. Model checks for regression: an innovation process approach. Annals of Statistics 26, 1916–1934.

Whang, Y.-J., 1998. A consisitent Kolmogorov test of parametric regression functions. Unpublished manuscript. Department of Economics, Ewha University.

Whang, Y.-J., Andrews, D.W.K., 1993. Tests of specification for parametric and semiparametric models. Journal of Econometrics 57, 277–318.

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

Wu, C.F.J., 1986. Jackknife, 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.