TESTING A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE WITH IDENTIFICATION THROUGH INSTRUMENTAL VARIABLES

by

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60208
USA

January 2004

ABSTRACT

This paper is concerned with inference about a function g that is identified by a conditional moment restriction involving instrumental variables. The paper presents the first test of the hypothesis that g belongs to a finite-dimensional parametric family against a nonparametric alternative. The test does not require nonparametric estimation of g and is not subject to the ill-posed inverse problem of nonparametric instrumental variables estimation. Under mild conditions, the test is consistent against any alternative model. Moreover, it has power exceeding the probability of rejecting a correct null hypothesis uniformly over a class of alternatives whose distance from the null hypothesis is $O(n^{-1/2})$, where n is the sample size.

Keywords: Hypothesis test, instrumental variables, specification testing, consistent testing

Part of this research was carried out while I was a visitor at the Centre for Microdata Methods and Practice, University College London. I thank Richard Blundell for many helpful discussions and comments. Research supported in part by NSF Grant SES 9910925.

TESTING A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE WITH IDENTIFICATION THROUGH INSTRUMENTAL VARIABLES

1. INTRODUCTION

Let Y be a scalar random variable, X and W be continuously distributed random scalars or vectors, and g be a function that is identified by the relation

$$(1.1) \qquad \boldsymbol{E}[Y - g(X) \mid W] = 0 .$$

In (1.1), Y is the dependent variable, X is a possibly endogenous explanatory variable, and W is an instrument for X. This paper presents, for the first time, a test of the null hypothesis that g in (1.1) belongs to a finite-dimensional parametric family against a nonparametric alternative hypothesis. Specifically, let Θ be a compact subset of $\mathbb{R}^d$ for some finite integer $d > 0$. The null hypothesis, H_0, is that

$$(1.2) \qquad g(x) = G(x, \theta)$$

for some $\theta \in \Theta$ and almost every x, where G is a known function. The alternative hypothesis, H_1, is that there is no $\theta \in \Theta$ such that (1.2) holds for almost every x. Under mild conditions, the test presented here is consistent against any alternative model. Moreover, in large samples the test has power exceeding the probability of rejecting a correct H_0 uniformly over a class of alternative models whose "distance" from H_0 is $O(n^{-1/2})$, where n is the sample size.

There has been much recent interest in nonparametric estimation of g in (1.1). See, for example, Newey, Powell and Vella (1999); Newey and Powell (2003); Darolles, Florens, and Renault (2002); Blundell, Chen, and Kristensen, (2003); and Hall and Horowitz (2003). However, methods for testing a parametric model of g against a nonparametric alternative do not yet exist. This paper presents the first such test. In contrast, there is a large literature on testing a parametric model of a conditional mean or quantile function against a nonparametric alternative.[1]

Testing is particularly important in (1.1) because it provides the only currently available form of inference about g that does not require g to be known up to a finite-dimensional parameter. Obtaining the asymptotic distribution of a nonparametric estimator of g is very difficult, and no existing estimator has a known asymptotic distribution. Nor is there a currently known method for obtaining a nonparametric confidence band for g. By contrast, the test statistic described in this paper has a relatively simple asymptotic distribution, and implementation of the test is not difficult.

The test developed here does not require nonparametric estimation of g and, therefore, is not affected by the ill-posed inverse problem of nonparametric instrumental variables estimation. Consequently, the "precision" of the test is greater than that of any nonparametric estimator of g. The rate of convergence in probability of a nonparametric estimator of g is always slower than $O_p(n^{-1/2})$ and, depending on the details of the probability distribution of (Y,X,W), may be slower than $O_p(n^{-\varepsilon})$ for any $\varepsilon > 0$ (Hall and Horowitz 2003). In contrast, the test described in this paper can detect a large class of nonparametric alternative models whose distance from the null-hypothesis model is $O(n^{-1/2})$. Nonparametric estimation and testing of conditional mean and median functions is another setting in which the rate of testing is faster than the fastest possible rate of estimation. See, for example, Guerre and Lavergne (2002) and Horowitz and Spokoiny (2001, 2002).

Section 2 of this paper presents the test for the special case in which X and W are scalars. The extension to multivariate X and W is in Section 3. Section 4 presents the results of a Monte Carlo investigation of the finite-sample performance of the test, and Section 5 presents an illustrative application of the test to real data. The proofs of theorems are in the appendix.

2. THE TEST WHEN X AND W ARE SCALARS

The assumption that X and W are scalars enables the main ideas of this paper to be presented with a minimum of notational and technical complexity. Rewrite (1.1) as

(2.1) $\quad Y = g(X) + U; \quad \boldsymbol{E}(U \mid W) = 0$,

where $U = Y - g(X)$. Assume that the support of (X,W) is contained within the unit square. This assumption can always be satisfied by carrying out a monotone transformation of (X,W). The data, $\{Y_i, X_i, W_i : i = 1,...,n\}$, are an independent random sample of (Y,X,W).

2.1 *The Test Statistic*

To develop the test statistic, let f_{XW} denote the probability density function of (X,W). Define the operator T on $L_2[0,1]$ by

$$T\nu(z) = \int t(x,z)\nu(x)dx,$$

where ν is any square integrable function and

$$t(x,z) = \int f_{XW}(x,w) f_{XW}(z,w)dw.$$

Assume that T is nonsingular. Consider two functions $G_1(x,\theta)$ and $G_2(x,\theta)$ to be equal if they differ only on a set of x values with Lebesgue measure 0. Then H_0 is equivalent to

(2.2) $$S(z) \equiv T[g - G(\cdot,\theta)](z) = 0$$

for some $\theta \in \Theta$ and almost every $z \in [0,1]$. H_1 is equivalent to the statement that there is no θ such that (2.2) holds. A test statistic can be based on a sample analog of

(2.3) $$\int_0^1 S(z)^2 dz .$$

To form the analog, let $\hat{f}_{XW}^{(-i)}$ denote a leave-observation-*i*-out kernel estimator of f_{XW}. That is, for a kernel function K and bandwidth h

$$\hat{f}_{XW}^{(-i)}(x,w) = \frac{1}{nh^2}\sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x - X_i}{h}\right) K\left(\frac{w - W_i}{h}\right).$$

Let $\hat{\theta}_n$ be an estimator of θ that is consistent under H_0. Then the sample analog of $S(z)$ is

(2.4) $$S_n(z) = n^{-1/2}\sum_{i=1}^{n}[Y_i - G(X_i,\hat{\theta}_n)]\hat{f}_{XW}^{(-i)}(z,W_i) .$$

The test statistic is

(2.5) $$\tau_n = \int_0^1 S_n^2(z) dz .$$

H_0 is rejected if τ_n is large.

2.2 *Regularity Conditions*

This section states the assumptions that are used to obtain the asymptotic properties of τ_n under the null and alternative hypotheses. The following additional notation is used. Let $\|(x_1,w_1) - (x_2,w_2)\|$ denote the Euclidean distance between the points (x_1,w_1) and (x_2,w_2) in $[0,1]^2$. Let $D_j f_{XW}$ denote any j'th partial or mixed partial derivative of f_{XW}. Set $D_0 f_{XW}(x,w) = f_{XW}(x,w)$. The assumptions are as follows.

1. (i) The support of (X,W) is contained in $[0,1]^2$. (ii) (X,W) has a probability density function $f_{XW}(x,w)$ with respect to Lebesgue measure. (iii) There is a constant $C_f < \infty$ such that $|D_j f_{XW}(x,w)| \leq C_f$ for all $(x,w) \in [0,1]^2$ and $j = 0,1,2$. (iv) $|D_2 f_{XW}(x_1,w_1) - D_2 f_{XW}(x_2,w_2)| \leq C_f \|(x_1,w_1) - (x_2,w_2)\|$ for any second derivative and any (x_1,w_1) and (x_2,w_2) in $[0,1]^2$. (v) The operator T is nonsingular.

2. (i) $\boldsymbol{E}(U \mid W=w)=0$ and $\boldsymbol{E}(U^2 \mid W=w) \le C_U$ for each $w \in [0,1]$ and some constant $C_U < \infty$. (ii) $|g(x)| \le C_g$ for some constant $C_g < \infty$ and all $x \in [0,1]$.

3. (i) As $n \to \infty$, $\theta_n \to^p \theta_0$ for some $\theta_0 \in \Theta$, a compact subset of $\mathbb{R}^d$. If H_0 is true, then $g(x) = G(x,\theta_0)$, $\theta_0 \in \text{int}(\Theta)$, and

(2.6) $$n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2}\sum_{i=1}^{n}\gamma(U_i, X_i, W_i, \theta_0) + o_p(1)$$

for some function γ taking values in $\mathbb{R}^d$ such that $\boldsymbol{E}\gamma(Y,X,W,\theta_0)=0$ and $Var[\gamma(Y,X,W,\theta_0)]$ is a finite, non-singular matrix.

4. (i) $|G(x,\theta)| \le C_G$ for all $x \in [0,1]$, all $\theta \in \Theta$, and some constant $C_G < \infty$. (ii) The first and second derivatives of $G(x,\theta)$ with respect to θ are bounded by C_G uniformly over $x \in [0,1]$ and $\theta \in \Theta$.

5. (i) The kernel function, K, is a symmetrical, twice continuously differentiable probability density function on $[-1,1]$. (ii) The bandwidth, h, satisfies $h = c_h n^{-1/6}$, where c_h is a constant and $0 < c_h < \infty$.

The representation (2.6) of $n^{1/2}(\hat{\theta}_n - \theta_0)$ holds, for example, if $\hat{\theta}_n$ is a generalized method of moments estimator

2.3 *The Asymptotic Distribution of the Test Statistic under the Null Hypothesis*

To obtain the asymptotic distribution of τ_n under H_0, define $G_\theta(x,\theta) = \partial G(x,\theta)/\partial\theta$, $\Gamma(z) = \boldsymbol{E}[G_\theta(X,\theta_0) f_{XW}(z,W)]$,

$$B_n(z) = n^{-1/2}\sum_{i=1}^{n}[U_i f_{XW}(z,W_i) - \Gamma(z)'\gamma(U_i,X_i,W_i,\theta_0)],$$

and $V(z_1,z_2) = \boldsymbol{E}[B_n(z_1)B_n(z_2)]$. Define the operator Ω on $L_2[0,1]$ by

$$(\Omega\psi)(z) = \int_0^1 V(z,x)\psi(x)dx .$$

Let $\{\omega_j : j=1,2,...\}$ denote the eigenvalues of Ω sorted so that $\omega_1 \ge \omega_2 \ge ... \ge 0$. Let $\{\chi_{1j}^2 : j=1,2,...\}$ denote independent random variables that are distributed as chi-square with one degree of freedom. The following theorem gives the asymptotic distribution of τ_n under H_0.

Theorem 1: Let H_0 be true. Then under assumptions 1-5,

$$\tau_n \to^d \sum_{j-1}^{\infty} \omega_j \chi_{1j}^2 .$$

2.4 *Obtaining the Critical Value*

The statistic τ_n is not asymptotically pivotal, so its asymptotic distribution cannot be tabulated. This section presents a method for obtaining an approximate asymptotic critical value for the τ_n test. The method is based on replacing the asymptotic distribution of τ_n with an approximate distribution. The difference between the true and approximate distributions can be made arbitrarily small under both the null hypothesis and alternatives. Moreover, the quantiles of the approximate distribution can be estimated consistently as $n \to \infty$. Accordingly, the proposed approximate $1-\alpha$ critical value of the τ_n test is a consistent estimator of the $1-\alpha$ quantile of the approximate distribution.

The approximate critical value is obtained under sampling from a pseudo-true model that coincides with (2.1) if H_0 is true and satisfies a version of $\boldsymbol{E}[Y - G(X,\theta_0) \mid W] = 0$ if H_0 is false. The critical value for the case of a false H_0 is used later to establish the properties of τ_n under H_1. The pseudo-true model is defined by

(2.7) $\quad \tilde{Y} = G(X,\theta) + \tilde{U}$,

where $\tilde{Y} = Y - \boldsymbol{E}[Y - G(X,\theta_0) \mid W]$, $\tilde{U} = \tilde{Y} - G(X,\theta_0)$, and θ_0 is the probability limit of $\hat{\theta}_n$. This model coincides with (2.1) when H_0 is true. Moreover, H_0 holds for the pseudo-true model in the sense that $\boldsymbol{E}[\tilde{Y} - G(X,\theta_0) \mid W] = 0$, regardless of whether H_0 holds for model (2.1).

To describe the approximation to the asymptotic distribution of τ_n, let $\{\tilde{\omega}_j : j = 1,2,...\}$ be the eigenvalues of the version of Ω (denoted $\tilde{\Omega}$) that is obtained by replacing model (2.1) with model (2.7). Order the $\tilde{\omega}_j$'s such that $\tilde{\omega}_1 \ge \tilde{\omega}_2 \ge ... \ge 0$. Then under sampling from (2.7), τ_n is asymptotically distributed as

$$\tilde{\tau} \equiv \sum_{j=1}^{\infty} \tilde{\omega}_j \chi_{1j}^2 .$$

Given any $\varepsilon > 0$, there is an integer $K_\varepsilon < \infty$ such that

$$0 < \boldsymbol{P}\left(\sum_{j=1}^{K_\varepsilon} \tilde{\omega}_j \chi_{1j}^2 \le t \right) - \boldsymbol{P}(\tilde{\tau} \le t) < \varepsilon .$$

uniformly over t. Define

$$\tilde{\tau}_{\varepsilon} = \sum_{j=1}^{K_{\varepsilon}} \tilde{\omega}_j \chi_{1j}^2 .$$

Let $z_{\varepsilon\alpha}$ denote the $1-\alpha$ quantile of the distribution of $\tilde{\tau}_{\varepsilon}$. Then $0 < \boldsymbol{P}(\tilde{\tau} > z_{\varepsilon\alpha}) - \alpha < \varepsilon$. Thus, using $z_{\varepsilon\alpha}$ to approximate the asymptotic $1-\alpha$ critical value of τ_n creates an arbitrarily small error in the probability that a correct null hypothesis is rejected. Similarly, use of the approximation creates an arbitrarily small change in the power of the τ_n test when the null hypothesis is false. However, the distribution of $\tilde{\tau}_{\varepsilon}$ is unknown because the eigenvalues $\tilde{\omega}_j$ are unknown. Accordingly, the approximate $1-\alpha$ critical value for the τ_n test is a consistent estimator of the $1-\alpha$ quantile of the distribution of $\tilde{\tau}_{\varepsilon}$. Specifically, let $\hat{\omega}_j$ $(j = 1, 2, ..., K_{\varepsilon})$ be a consistent estimator of $\tilde{\omega}_j$ under sampling from (2.7). Then the approximate critical value of τ_n is the $1-\alpha$ quantile of the distribution of

$$\hat{\tau}_n = \sum_{j=1}^{K_{\varepsilon}} \hat{\omega}_j \chi_{1j}^2 .$$

This quantile, which will be denoted $\hat{z}_{\varepsilon\alpha}$, can be estimated with arbitrary accuracy by simulation.[2]

The remainder of this section describes how to obtain the estimated eigenvalues $\{\hat{\omega}_j\}$. For this purpose, it is assumed that $\hat{\theta}_n$ satisfies the estimating equation

$$n^{-1} \sum_{i=1}^{n} \tilde{W}_i [\tilde{Y}_i - G(X_i, \hat{\theta}_n)] = 0 ,$$

where $\tilde{W}_i = H(W_i)$ for some known function $H : \mathbb{R} \to \mathbb{R}^d$ whose components are linearly independent. For example, $\tilde{W}_i$ might be a vector whose components are powers of W_i. By an application of the delta method,

$$\gamma(\tilde{U}, X, W, \theta_0) = Q^{-1} \tilde{W} \tilde{U} ,$$

where $\tilde{W} = H(W)$, $Q = \boldsymbol{E}[\tilde{W} G_{\theta}(X, \theta_0)']$, and Q is assumed to be non-singular. Some algebra now shows that

$$V(z_1, z_2) = \boldsymbol{E}\{[f_{XW}(z_1, W) - \Gamma(z_1)' Q^{-1} \tilde{W}] \tilde{U}^2 [f_{XW}(z_2, W) - \Gamma(z_2)' Q^{-1} \tilde{W}]\} . \tag{2.8}$$

A consistent estimator of $V(z_1, z_2)$ can be obtained by replacing the unknown quantities on the right-hand side of (2.8) with estimators. Let $\hat{f}_{XW}$ a kernel estimator of f_{XW} with bandwidth h. Define

$$\hat{Q} = n^{-1}\sum_{i=1}^{n}\tilde{W}_i G_\theta(X_i,\hat{\theta}_n)'$$

and

$$\hat{\Gamma}(z) = n^{-1}\sum_{i=1}^{n}\hat{f}_{XW}(z,W_i)G_\theta(X_i,\hat{\theta}_n).$$

Let $\hat{q}^{(-i)}(w)$ be the leave-observation i-out kernel estimator

$$\hat{q}^{(-i)}(w) = \sum_{\substack{j=1\\ j\neq i}}^{n}\kappa_h(w-W_j)[Y_j - G(X_j,\hat{\theta}_n],$$

where

$$\kappa_h(w-W_j) = K\left(\frac{w-W_j}{h}\right)\left[\sum_{\substack{k=1\\ k\neq i}}^{n}K\left(\frac{w-W_k}{h}\right)\right]^{-1}$$

and $h \propto n^{-1/5}$ is the bandwidth. The $\tilde{U}_i$'s are estimated by residuals of model (2.7). These are

$$\hat{U}_i = Y_i - G(X_i,\hat{\theta}_n) - \hat{q}^{(-i)}(W_i).$$

Then $V(z_1,z_2)$ is estimated consistently by

$$\hat{V}(z_1,z_2) = n^{-1}\sum_{i=1}^{n}\{[\hat{f}_{XW}(z_1,W_i) - \hat{\Gamma}(z_1)'\hat{Q}^{-1}\tilde{W}_i]\hat{U}_i^2[\hat{f}_{XW}(z_2,W_i) - \hat{\Gamma}(z_2)'\hat{Q}^{-1}\tilde{W}_i].$$

Define $\hat{\Omega}$ to be the integral operator whose kernel is $\hat{V}(z_1,z_2)$. The $\hat{\omega}_j$'s are the eigenvalues of $\hat{\Omega}$.

<u>Theorem 2</u>: Let assumptions 1-5 hold. Then as $n \to \infty$, (i) $\sup_{1\leq j\leq K_\varepsilon}|\hat{\omega}_j - \tilde{\omega}_j| = O[(\log n)/(nh^2)^{1/2}]$ almost surely and (ii) $\hat{z}_{\varepsilon\alpha} \to^p z_{\varepsilon\alpha}$.

To obtain an accurate numerical approximation to the $\hat{\omega}_j$'s, let $\hat{F}(z)$ denote the $n\times 1$ vector whose i'th component is $\hat{f}_{XW}(z,W_i)$, $\hat{G}_\theta$ denote the $n\times d$ matrix whose (i,j) element is $G_\theta(X_i,\hat{\theta}_n)$, Υ denote the $n\times n$ diagonal matrix whose (i,i) element is $\hat{U}_i^2$, and $\tilde{W}$ denote the $n\times d$ matrix $(\tilde{W}_1',...,\tilde{W}_n')'$. Finally, define the matrix $M = I_n - n^{-1}\hat{G}_\theta\hat{Q}^{-1}\tilde{W}'$, where I_n is the $n\times n$ identity matrix. Then

$$\hat{V}(z_1,z_2) = n^{-1}\hat{F}(z_1)'M\Upsilon M'\hat{F}(z_2).$$

The computation of the $\hat{\omega}_j$'s can now be reduced to finding the eigenvalues of a finite-dimensional matrix. To this end, let $\{\phi_j : j = 1,2,...\}$ be an orthonormal basis for $L_2[0,1]$. Then

$$\hat{f}_{XW}(z,W) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \hat{d}_{jk} \phi_j(z) \phi_k(W),$$

where

$$\hat{d}_{jk} = \int_0^1 dx \int_0^1 dw \hat{f}_{XW}(x,w) \phi_j(x) \phi_k(w).$$

Approximate $\hat{f}_{XW}$ by the finite sum

$$\hat{\hat{f}}_{XW}(z,W) = \sum_{j=1}^{L} \sum_{k=1}^{L} \hat{d}_{jk} \phi_j(z) \phi_k(W)$$

for some integer $L < \infty$. Since $\hat{f}_{XW}$ is a known function, L can be chosen to make $\hat{\hat{f}}_{XW}$ approximate $\hat{f}_{XW}$ with any desired accuracy. Let $\phi(z)$ denote the $L \times 1$ vector whose j'th component is $\phi_j(z)$. Let Φ be the $L \times n$ matrix whose (j,k) component is $\phi_j(W_k)$. Let D be the $L \times L$ matrix $\{d_{jk}\}$. Then $\hat{V}(z_1, z_2)$ is approximated by

$$\hat{\hat{V}}(z_1, z_2) = n^{-1} \phi(z_1)' D \Phi M \Upsilon M' \Phi' D' \phi(z_2).$$

The eigenvalues of $\hat{\Omega}$ are approximated by those of the $L \times L$ matrix $D \Phi M \Upsilon M' \Phi' D'$.

2.5 *Consistency of the Test against a Fixed Alternative Model*

In this section, it is assumed that H_0 is false. That is, there is no $\theta \in \Theta$ such that $g(x) = G(x,\theta)$ for almost every x. Let θ_0 denote the probability limit of $\hat{\theta}_n$. Define $q(x) = g(x) - G(x,\theta_0)$. Let $\tilde{z}_\alpha$ denote the $1-\alpha$ quantile of the distribution of τ_n under sampling from the pseudo-true model (2.7). Let $\hat{z}_{\varepsilon\alpha}$ denote the $1-\alpha$ quantile of $\hat{\tau}_n$. The following theorem establishes consistency of the τ_n test against a fixed alternative hypothesis.

Theorem 3: Suppose that

$$\int_0^1 [(Tq)(z)]^2 \, dz > 0$$

Let assumptions 1-5 hold. Then for any α such that $0 < \alpha < 1$,

(2.9) $$\lim_{n \to \infty} \boldsymbol{P}(\tau_n > \tilde{z}_\alpha) = 1$$

and

(2.10) $\lim_{n\to\infty} \boldsymbol{P}(\tau_n > \hat{z}_\alpha) = 1$.

Because T is nonsingular, the τ_n test is consistent against any alternative that differs from $G(x,\theta_0)$ on a set of x values whose Lebesgue measure exceeds zero.

2.6 *Asymptotic Distribution under Local Alternatives*

This section obtains the asymptotic distribution of τ_n under the sequence of local alternative hypotheses

$$g_n(x) = G(x,\theta_0) + n^{-1/2}\Delta(x),$$

where Δ is a bounded function on $[0,1]$ and $\theta_0 \in \text{int}(\Theta)$. Under this sequence of local alternatives, the data are generated by the model

(2.11) $Y = G(X,\theta_0) + n^{-1/2}\Delta(X) + U$.

The following additional notation is used to state the result. Let $\{\psi_j : j = 1,2,...\}$ denote the orthonormal eigenvectors of Ω. Define $\mu(z) = (T\Delta)(z)$ and

$$\mu_j = \int_0^1 \mu(z)\psi_j(z)dz.$$

Let $\{\chi^2_{1j}(\mu_j) : j = 1,2,...\}$ denote independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $\{\mu_j\}$. The following theorem states the result.

<u>Theorem 4</u>: Let assumptions 1-5 hold. Under the sequence of local alternatives (2.11),

$$\tau_n \to^d \sum_{j-1}^{\infty} \omega_j \chi^2_{1j}(\mu_j),$$

where the ω_j 's are the eigenvalues of the operator Ω defined in (2.6).

Let z_α denote the $1-\alpha$ quantile of the distribution of $\sum_{j=1}^{\infty} \omega_j \chi^2_{1j}(\mu_j)$. Let $\hat{z}_{\varepsilon\alpha}$ denote the estimated approximate α -level critical value defined in Section 2.2. Then it follows from Theorems 2 and 4 that for any $\varepsilon > 0$ and all sufficiently large n,

$$\limsup_{n\to\infty} |\boldsymbol{P}(\tau_n > \hat{z}_{\varepsilon\alpha}) - \boldsymbol{P}(\tau_n > z_\alpha)| \le \varepsilon.$$

2.7 *Uniform Consistency*

This section shows that for any $\varepsilon > 0$, the τ_n test rejects H_0 with probability exceeding $1-\varepsilon$ uniformly over a class of alternative models whose distance from the null hypothesis is $O(n^{-1/2})$. The following additional notation is used. Let θ_g be the probability limit of $\hat{\theta}_n$ under the hypothesis (not necessarily true) that $g(x) = G(x,\theta)$ for some $\theta \in \Theta$ and a given function G. Define $q_g(x) = g(x) - G(x,\theta_g)$. Let h denote the bandwidth in $f_{XW}^{(-i)}$. For each $n = 1,2.,,,$ and $C > 0$ define $\mathcal{F}_{nc}$ as a set of functions g such that: (i) $|g(x)| \le C_g$, for all $x \in [0,1]$ and some constant $C_g < \infty$; (ii) $\theta_g \in \Theta$; (iii) for each $\varepsilon > 0$ there is a $M_\varepsilon < \infty$ such that $\boldsymbol{P}\left[n^{1/2} \sup_{x\in[0,1], g\in\mathcal{F}_{nc}} \left| G(x,\hat{\theta}_n) - G(x,\theta_g) \right| > M_\varepsilon \right] < \varepsilon$; (iv) $\|Tq_g\| \ge n^{-1/2}C$, where $\|\cdot\|$ denotes the L_2 norm; and (v) $h^2 \|q_g\| / \|Tq_g\| = o(1)$ as $n \to \infty$. $\mathcal{F}_{nc}$ is a set of functions whose distance from H_0 shrinks to zero at the rate $n^{-1/2}$. That is, $\mathcal{F}_{nc}$ includes functions such that $\|q_g\| = O(n^{-1/2})$. Condition (v) rules out alternatives that depend on x only through sequences of eigenvectors of T whose eigenvalues converge to 0 too rapidly. For example, let $\{\lambda_j, \phi_j : j = 1,2,...\}$ denote the eigenvalues and eigenvectors of T ordered so that $\lambda_1 \ge \lambda_2 \ge ... > 0$. Suppose that $G(x,\theta) = \theta\phi_1(x)$, $g(x) = \phi_1(x) + \phi_n(x)$, and $\tilde{W} = \phi_1(W)$. Then $h^2 \|q_g\| / \|Tq_g\| = h^2/\lambda_n$. Because $h \propto n^{-1/6}$, condition (v) is violated if $\lambda_n = o(n^{-1/3})$. The practical significance of condition (v) is that the τ_n test has relatively low power against alternatives that differ from the null hypothesis only through eigenvectors of T with very small eigenvalues.

The following theorem states the result of this section.

Theorem 5: Let Assumptions 1, 2, 4, and 5 hold. Then given any $\varepsilon > 0$, any α such that $0 < \alpha < 1$, and any sufficiently large (but finite) C,

(2.12) $$\liminf_{n\to\infty} \inf_{\mathcal{F}_{nc}} \boldsymbol{P}(\tau_n > z_\alpha) \ge 1 - \varepsilon$$

and

(2.13) $$\liminf_{n\to\infty} \inf_{\mathcal{F}_{nc}} \boldsymbol{P}(\tau_n > \hat{z}_{\varepsilon\alpha}) \ge 1 - 2\varepsilon .$$

3. MULTIVARIATE GENERALIZATION

This section generalizes the τ_n test to a multivariate version of (1.1) and (2.1) in which some of the explanatory variables may be exogenous. The model is

(3.1) $Y = g(X,Z) + U;\ \boldsymbol{E}(U \mid Z,W) = 0$,

where Y and U are scalar random variables, X and W are random variables whose supports are contained in $[0,1]^p$ ($p \geq 1$), and Z is a random variable whose support is contained in $[0,1]^r$ ($r \geq 0$). If $r = 0$, then Z is not a variable of the model. In (3.1), X and Z, respectively, are endogenous and exogenous explanatory variables. W is an instrument for X. The inferential problem is to test the null hypothesis, H_0, that

(3.2) $g(x,z) = G(x,z,\theta)$

for some unknown $\theta \in \Theta$, known function G, and almost every $(x,z) \in [0,1]^{p+r}$. The alternative hypothesis, H_1 is that there is no $\theta \in \Theta$ such that (3.2) holds for almost every $(x,z) \in [0,1]^{p+r}$. The data, $\{Y_i, X_i, Z_i, W_i : i = 1,...,n\}$, are a simple random sample of (Y,X,Z,W).

3.1 *The Test Statistic*

To form the test statistic, let f_{XZW} denote the probability density function of (X,Z,W), and let f_Z denote the probability density function of Z. Let ν be any function in $L_2[0,1]^{p+r}$. For each $z \in [0,1]^r$ define the operator T_z on $L_2[0,1]^p$ by

$$T_Z \nu(x,z) = \int t_z(\xi, x)\nu(\xi, z)d\xi,$$

where for each $(x_1, x_2) \in [0,1]^{2p}$,

$$t_z(x_1, x_2) = \int f_{XZW}(x_1, z, w) f_{XZW}(x_2, z, w)dw.$$

Assume that T_z is nonsingular for each $z \in [0,1]^r$. Then H_0 is equivalent to

(3.3) $\tilde{S}(x,z) \equiv T_z[g(\cdot,\cdot) - G(\cdot,\cdot,\theta)](x,z) = 0$

for some $\theta \in \Theta$ and almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that there is no $\theta \in \Theta$ such that (3.3) holds almost every $(x,z) \in [0,1]^{p+r}$. A test statistic can be based on a sample analog of

$$\int \tilde{S}(x,z)^2 dxdz,$$

but the resulting rate of testing is slower than $n^{-1/2}$ if $r > 0$. A rate of $n^{-1/2}$ can be achieved, though at the cost of uniform consistency over a smaller class of alternatives, by carrying out an additional smoothing step. To this end, let $\ell(z_1, z_2)$ denote the kernel of a nonsingular integral operator, L, on $L_2[0,1]^r$. That is, the operator L defined by

$$Lv(z) = \int \ell(\zeta, z)v(\zeta)d\zeta$$

is nonsingular. Define the operator T_M on $L_2[0,1]^{p+r}$ by $T_M v(x,z) = LT_z v(x,z)$. Then T_M is non-singular. H_0 is equivalent to

(3.4) $$S_M(x,z) \equiv T_M[g(\cdot,\cdot) - G(\cdot,\cdot,\theta)](x,z) = 0$$

for some $\theta \in \Theta$ and almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that there is no $\theta \in \Theta$ such that (3.5) holds. The test statistic is based on a sample analog of

$$\int S_M(x,z)^2 \, dxdz .$$

To form the analog, let $\hat{f}_{XZW}^{(-i)}$ denote a leave-observation-*i*-out kernel estimator of f_{XZW}. That is, for $V_i \equiv (X_i, Z_i, W_i)$ and κ a kernel function of a $2p+r$-dimensional argument,

$$\hat{f}_{XZW}^{(-i)}(v) = \frac{1}{nh^{2p+r}} \sum_{\substack{j=1 \\ j \neq i}}^{n} \kappa\left(\frac{v - V_i}{h}\right),$$

where h is the bandwidth. Let $\hat{\theta}_n$ be an estimator of θ. The sample analog of $S_M(x,z)$ is

(3.5) $$S_{Mn}(x,z) = n^{-1/2} \sum_{i=1}^{n} [Y_i - G(X_i, Z_i, \hat{\theta}_n)] \hat{f}_{XZW}^{(-i)}(x, Z_i, W_i) \ell(Z_i, z) .$$

The test statistic is

(3.6) $$\tau_{Mn} = \int S_{Mn}^2(x,z)dxdz$$

H_0 is rejected if τ_{Mn} is large.

3.2 *Regularity Conditions*

This section states the assumptions that are used to obtain the asymptotic properties of τ_{Mn} under the null and alternative hypotheses. Let $\|(x_1, z_1, w_1) - (x_2, z_2, w_2)\|$ denote the Euclidean distance between (x_1, z_1, w_1) and (x_2, z_2, w_2).

M1. (i) The support of (X,Z,W) is contained in $[0,1]^{2p+r}$. (ii) (X,Z,W) has a probability density function f_{XZW} with respect to Lebesgue measure. (iii) There is a constant $C_f < \infty$ such that $|D_j f_{XZW}(x,z,w)| \leq C_f$ for all $(x,z,w) \in [0,1]^{2p+r}$ and $j = 0,1,2$. (iv) $|D_2 f_{XZW}(x_1, z_1, w_1) - D_2 f_{XZW}(x_2, z_2, w_2)| \leq C_f \|(x_1, z_1, w_1) - (x_2, z_2, w_2)\|$ for any second derivative and any (x_1, z_1, w_1) and (x_2, z_2, w_2) in $[0,1]^{2p+r}$. (v) The operator T_z is nonsingular for almost every $z \in [0,1]^r$.

M2. (i) $\boldsymbol{E}(U \mid Z = z, W = w) = 0$ and $\boldsymbol{E}(U^2 \mid Z = z, W = w) \le C_U$ for each $(z, w) \in [0,1]^{p+r}$ and some constant $C_U < \infty$. (ii) $|g(x,z)| \le C_g$ for some constant $C_g < \infty$ and all $(x,z) \in [0,1]^{p+r}$.

M3. (i) As $n \to \infty$, $\theta_n \to^p \theta_0$ for some $\theta_0 \in \Theta$, a compact subset of $\mathbb{R}^d$. If H$_0$ is true, then $g(x,z) = G(x,z,\theta_0)$, $\theta_0 \in \text{int}(\Theta)$, and

$$n^{-1/2}(\hat{\theta}_n - \theta_0) = n^{1/2} \sum_{i=1}^{n} \gamma(U_i, X_i, Z_i, W_i, \theta_0) + o_p(1)$$

for some function γ taking values in $\mathbb{R}^d$ such that $\boldsymbol{E}\gamma(U, X, Z, W, \theta_0) = 0$ and $Var[\gamma(U, X, Z, W, \theta_0)]$ is a finite, non-singular matrix.

M4. (i) $|G(x,z,\theta)| \le C_G$ for all $(x,z) \in [0,1]^{p+r}$, all $\theta \in \Theta$, and some constant $C_G < \infty$. (ii) The first and second derivatives of $G(x,z,\theta)$ with respect to θ are bounded by C_G uniformly over $(x,z) \in [0,1]^{p+r}$ and $\theta \in \Theta$.

M5. (i) The kernel function used to estimate f_{XZW} has the form $\kappa(v) = \prod_{j=1}^{2p+r} K(v_j)$, where v_j is the j'th component of v and K is a symmetrical, twice continuously differentiable probability density function on $[-1,1]$. (ii) The bandwidth, h, satisfies $h = c_h n^{-1/(2p+r+4)}$, where c_h is a constant and $0 < c_h < \infty$. (iii) The operator L is nonsingular.

3.2 *Asymptotic Distribution under the Null Hypothesis*

The asymptotic distributional properties of τ_{Mn} are similar to those of τ_n. To state the asymptotic distribution of τ_{Mn} under H$_0$, define $G_\theta(x,z,\theta) = \partial G(x,z,\theta)/\partial\theta$, $\Gamma(x,z) = \boldsymbol{E}[G_\theta(X, Z, \theta_0) f_{XZW}(x, Z, W)\ell(Z, z)]$,

$$B_{Mn}(x,z) = n^{-1/2} \sum_{i=1}^{n} [U_i f_{XZW}(x, Z_i, W_i)\ell(Z_i, z) - \Gamma(x,z)'\gamma(U_i, X_i, Z_i, W_i, \theta_0)],$$

and

$$V_M(x_1, z_1; x_2, z_2) = \boldsymbol{E}[B_{Mn}(x_1, z_1) B_{Mn}(x_2, z_2)].$$

Define the operator Ω_M on $L_2[0,1]^{q+r}$ by

(3.7) $(\Omega_M \nu)(x,z) = \int_0^1 V_M(x,z;\xi,\zeta)\nu(\xi,\zeta)d\xi d\zeta$.

Let $\{\omega_{Mj}, \psi_{Mj} : j = 1,2,...\}$ denote the eigenvalues and orthonormal eigenvectors of Ω_M sorted so that $\omega_{M1} \geq \omega_{M2} \geq ... \geq 0$. Let $\{\chi^2_{1j} : j = 1,2,...\}$ denote independent random variables that are distributed as chi-square with one degree of freedom. The asymptotic distribution of τ_{Mn} under H_0 is given by the following theorem.

Theorem 6: If H_0 is true and assumptions M1-M5 hold, then

$$\tau_{Mn} \to^d \sum_{j-1}^{\infty} \omega_{Mj} \chi^2_{1j} .$$

To obtain an approximate critical value for the τ_{Mn} test, define the pseudo-true model

(3.8) $\tilde{Y} = G(X,Z,\theta) + \tilde{U}$,

where $\tilde{Y} = Y - \boldsymbol{E}[Y - G(X,Z,\theta_0) \mid Z,W]$, $\tilde{U} = \tilde{Y} - G(X,Z,\theta_0)$, and θ_0 is the probability limit of $\hat{\theta}_n$. Let $\{\tilde{\omega}_{Mj} : j = 1,2,...\}$ be the eigenvalues of the version of Ω_M that is obtained by replacing model (3.1) with model (3.8). It follows from Theorem 6 that under sampling from (3.8), τ_{Mn} is asymptotically distributed as

$$\tilde{\tau}_M \equiv \sum_{j=1}^{\infty} \tilde{\omega}_{Mj} \chi^2_{1j} .$$

Let $z_{M\alpha}$ denote the $1-\alpha$ quantile of this distribution. The method for approximating this quantile in an application is similar to the method proposed for τ_n. Given any $\varepsilon > 0$, there is an integer $K_\varepsilon < \infty$ such that

$$0 < \boldsymbol{P}\left(\sum_{j=1}^{K_\varepsilon} \tilde{\omega}_{Mj} \chi^2_{1j} \leq t \right) - \boldsymbol{P}(\tilde{\tau}_M \leq t) < \varepsilon$$

uniformly over t. Define

$$\tilde{\tau}_{M\varepsilon} = \sum_{j=1}^{K_\varepsilon} \tilde{\omega}_{Mj} \chi^2_{1j} .$$

Let $z_{M\varepsilon\alpha}$ denote the $1-\alpha$ quantile of the distribution of $\tilde{\tau}_{M\varepsilon}$. Then using $z_{M\varepsilon\alpha}$ to approximate the asymptotic $1-\alpha$ critical value of τ_{Mn} creates an arbitrarily small error in the probability that a correct null hypothesis is rejected. The proposed approximate $1-\alpha$ critical value for the τ_{Mn} test is a consistent estimator of the $1-\alpha$ quantile of the distribution of $\tilde{\tau}_{M\varepsilon}$. Specifically, let

$\hat{\omega}_{Mj}$ $(j=1,2,...,K_\varepsilon)$ be the estimator of $\tilde{\omega}_{Mj}$ under sampling from (3.8) that is described below. Then the approximate critical value of τ_n, $\hat{z}_{M\varepsilon\alpha}$, is the $1-\alpha$ quantile of the distribution of

$$\hat{\tau}_{Mn} = \sum_{j=1}^{K_\varepsilon} \hat{\omega}_{Mj} \chi^2_{1j} .$$

The estimator of $\tilde{\omega}_{Mj}$ is the multivariate generalization of the estimator $\hat{\omega}_j$ for the bivariate model (2.1). Define $\tilde{W}_i = [H(W_i)', Z_i']'$, where H is a known, vector-valued function whose components are linearly independent, and $c_\theta \equiv \dim H + r \geq d$. Assume that $\hat{\theta}_n$ is the GMM estimator

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \left\{ \sum_{i=1}^{n} [Y_i - G(X_i, Z_i, \theta)] \tilde{W}_i' \right\} A_n \left\{ \sum_{i=1}^{n} [Y_i - G(X_i, Z_i, \theta)] \tilde{W}_i \right\},$$

where $\{A_n\}$ is a sequence of possibly stochastic $c_\theta \times c_\theta$ weight matrices converging in probability to a non-stochastic limit matrix A. Define the $c_\theta \times d$ matrix $D = \boldsymbol{E}[\tilde{W} G_\theta (X, Z, \theta_0)']$ and the $d \times c_\theta$ matrix $\tilde{\gamma}_M = (D'AD)^{-1} D'A$. Then standard calculations for GMM estimators show that

$$\gamma(U_i, X_i, Z_i, W_i, \theta_0) = \tilde{\gamma}_M \tilde{W}_i \tilde{U}_i .$$

Therefore,

$$V_M(x_1, z_1; x_2, z_2) = \boldsymbol{E} \left\{ n^{-1} \sum_{i=1}^{n} [f_{XZW}(x_1, Z_i, W_i) \ell(Z_i, z_1) - \Gamma(x_1, z_1)' \tilde{\gamma}_M \tilde{W}_i] \tilde{U}_i^2 \right.$$

$$\left. \times [f_{XZW}(x_2, Z_i, W_i) \ell(Z_i, z_2) - \Gamma(x_2, z_2)' \tilde{\gamma}_M \tilde{W}_i] \right\}.$$

To estimate V_M, define $\hat{D} = n^{-1} \sum_{i=1}^{n} \tilde{W}_i G_\theta (X, Z, \hat{\theta}_n)'$, $\hat{\gamma}_M = (\hat{D}' A_n \hat{D})^{-1} \hat{D}' A_n$, and

$$\hat{\Gamma}(x, z) = n^{-1} \sum_{i=1}^{n} G_\theta (X_i, Z_i, \hat{\theta}_n) \hat{f}_{XZW}(x, Z_i, W_i) \ell(Z_i, z),$$

where $\hat{f}_{XZW}$ is a kernel estimator of f_{XZW}. Also define $\hat{U}_i = Y_i - G(X_i, Z_i, \hat{\theta}_n) - \hat{q}_M^{(-i)}(Z_i, W_i)$, where $\hat{q}_M^{(-i)}(z, w)$ is the leave-observation i-out kernel regression estimator of $Y - G(X, Z, \hat{\theta}_n)$ on (Z, W). Then $V_M(x_1, z_1; x_2, z_2)$ is estimated consistently by

$$\hat{V}_M(x_1, z_1; x_2, z_2) = \left\{ n^{-1} \sum_{i=1}^{n} [\hat{f}_{XZW}(x_1, Z_i, W_i) \ell(Z_i, z_1) - \hat{\Gamma}(x_1, z_1)' \hat{\gamma}_M \tilde{W}_i] \hat{U}_i^2 \right.$$

$$\left. \times [\hat{f}_{XZW}(x_2, Z_i, W_i) \ell(Z_i, z_2) - \hat{\Gamma}(x_2, z_2)' \hat{\gamma}_M \tilde{W}_i] \right\}.$$

Let $\hat{\Omega}_M$ be the integral operator whose kernel is $\hat{V}_M(x_1, z_1; x_2, z_2)$. Then $\hat{\omega}_{Mj}$ is the j'th eigenvalue of $\hat{\Omega}_M$. The multivariate analog of Theorem 2 is:

Theorem 7: Let assumptions M1-M5 hold. Then as $n \to \infty$, (i) $\sup_{1 \le j \le K_\varepsilon} |\hat{\omega}_{Mj} - \tilde{\omega}_{Mj}| = O[(\log n)/(nh^{2p+r})^{1/2}]$ almost surely and (ii) $\hat{z}_{M\varepsilon\alpha} \to^p z_{M\varepsilon\alpha}$.

3.3 *Consistency against a Fixed Alternative Model*

Suppose that H_0 is false, meaning that there is no $\theta \in \Theta$ such that $g(x,z) = G(x,z,\theta)$ for almost every (x,z). Define $q(x,z) = g(x,z) - G(x,z,\theta_0)$. The following theorem establishes consistency of the τ_{Mn} test against a fixed alternative hypothesis.

Theorem 8: Let assumptions M1-M5 hold. Suppose that H_0 is false and that $\int [(T_M q)(x,z)]^2 dxdz > 0$. Then for any α such that $0 < \alpha < 1$,

$$\lim_{n \to \infty} \boldsymbol{P}(\tau_{Mn} > z_{M\alpha}) = 1 .$$

and

$$\lim_{n \to \infty} \boldsymbol{P}(\tau_{Mn} > \hat{z}_{M\varepsilon\alpha}) = 1 .$$

Because T_M is nonsingular, the τ_{Mn} test is consistent against any alternative that differs from $G(x,z,\theta_0)$ on a set of (x,z) values whose Lebesgue measure exceeds zero.

3.4 *Asymptotic Distribution under Local Alternatives*

This section obtains the asymptotic distribution of τ_{Mn} under the sequence of local alternative hypotheses

$$g_n(X,Z) = G(X,Z,\theta_0) + n^{-1/2}\Delta(X,Z) ,$$

where Δ is a bounded function on $[0,1]^{p+r}$. Under this sequence of local alternatives, the data are generated by the model

$$Y = G(X,Z,\theta_0) + n^{-1/2}\Delta(X,Z) + U . \tag{3.9}$$

Define

$$\mu_{Mj} = \int (T_M \Delta)(x,z) \psi_{Mj}(x,z) dxdz .$$

Let $\{\chi_{1j}^2(\mu_{Mj}): j = 1,2,...\}$ denote independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $\{\mu_{Mj}\}$. The following theorem states the result.

Theorem 9: Let assumptions M1-M5 hold. Under the sequence of alternatives (3.9),

$$\tau_{Mn} \to^d \sum_{j-1}^{\infty} \omega_{Mj} \chi_{1j}^2(\mu_{Mj}).$$

Let z_α denote the $1-\alpha$ quantile of the distribution of $\sum_{j=1}^{\infty} \omega_{Mj} \chi_{1j}^2(\mu_{Mj})$. Let $\hat{z}_{M\varepsilon\alpha}$ denote the estimated approximate α -level critical value pf τ_{Mn}. Then it follows from Theorems 7 and 9 that for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} | \boldsymbol{P}(\tau_{Mn} > \hat{z}_{M\varepsilon\alpha}) - \boldsymbol{P}(\tau_{Mn} > z_{M\alpha}) | \le \varepsilon .$$

3.5 *Uniform Consistency*

The multivariate version of $\mathcal{F}_{nc}$ is denoted $\mathcal{F}_{Mnc}$, and is defined as follows. As before, let θ_g be the probability limit of $\hat{\theta}_n$ under the hypothesis that $g(x,z) = G(x,z,\theta)$ for some $\theta \in \Theta$ and a given function G. Define $q_{Mg}(x,z) = g(x,z) - G(x,z,\theta_g)$. For each $n = 1,2.,,,$ and $C > 0$ define $\mathcal{F}_{Mnc}$ as a set of functions g such that: (i) $|g(x,z)| \le C_g$ for all $(x,z) \in [0,1]^{p+r}$ and some constant $C_g < \infty$; (ii) $\theta_g \in \Theta$; (iii) for each $\varepsilon > 0$ there is a $M_\varepsilon < \infty$ such that $\boldsymbol{P}\left[n^{1/2} \sup_{x,z\in[0,1]^{p+r}, g\in\mathcal{F}_{Mnc}} \left| G(x,z,\hat{\theta}_n) - G(x,z,\theta_g) \right| > M_\varepsilon \right] < \varepsilon$; (iv) $\left\| T_M q_{Mg} \right\| \ge n^{-1/2} C$; and (v) $h^2 \left\| q_{Mg} \right\| / \left\| T_M q_{Mg} \right\| = o(1)$ as $n \to \infty$. The following theorem states the multivariate uniform consistency result.

Theorem 10: Let assumptions M1, M2, M4, and M5 hold. Then given any $\varepsilon > 0$, α such that $0 < \alpha < 1$, and any sufficiently large but finite constant C,

$$\lim_{n\to\infty} \inf_{\mathcal{F}_{Mnc}} \boldsymbol{P}(\tau_{Mn} > z_{M\alpha}) \ge 1 - \varepsilon .$$

and

$$\lim_{n\to\infty} \inf_{\mathcal{F}_{Mnc}} \boldsymbol{P}(\tau_{Mn} > \hat{z}_{M\varepsilon\alpha}) \ge 1 - 2\varepsilon .$$

4. MONTE CARLO EXPERIMENTS

This section reports the results of a Monte Carlo investigation of the finite-sample performance of the τ_n test. The experiments consist of testing the hypothesis that

(4.1) $g(x) = \theta_0 + \theta_1 x$.

The alternative hypotheses are that g is quadratic,

(4.2) $g(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

and $g(x)$ is cubic,

(4.3) $g(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$.

To provide a basis for judging whether the power of the τ_n test is high or low, this section also reports the results of an asymptotic t test of the hypothesis $\theta_2 = 0$. The t test is an example of an *ad hoc* test that might be used in applied research. In all experiments, $\theta_0 = 0$ and $\theta_1 = 0.5$. In experiments where (4.2) is the correct model, $\theta_2 = -0.5$. In experiments where (4.3) is the correct model, $\theta_2 = -1$ and $\theta_3 = 1$. Realizations of (X, W) were generated by $X = \Phi(\xi)$, $W = \Phi(\zeta)$, where Φ is the cumulative normal distribution function, $\zeta \sim N(0,1)$, $\xi = \rho\zeta + (1-\rho^2)^{1/2}\varepsilon$, $\varepsilon \sim N(0,1)$, and ρ is a constant parameter whose value varies among experiments. Realizations of Y were generated from $Y = g(x) + \sigma_U U$, where $U = \eta\varepsilon + (1-\eta^2)^{1/2}\nu$, $\nu \sim N(0,1)$, $\sigma_U = 0.2$, and η is a constant parameter whose value varies among experiments. The instruments used to estimate (4.1), (4.2), and (4.3), respectively, are $(1, W)$, $(1, W, W^2)$, and $(1, W, W^2, W^3)$. The bandwidth h used to estimate f_{XW} was selected by cross-validation. The kernel is $K(v) = (15/16)(1-v^2)^2 I(|v| \le 1)$, where I is the indicator function. The asymptotic critical value was estimated by setting $K_\varepsilon = 25$. The results of the experiments are not sensitive to the choice of K_ε. The experiments use a sample size of $n = 500$ and the nominal 0.05 level. There are 1000 Monte Carlo replications in each experiment.

The results of the experiments are shown in Table 1. The differences between the nominal and empirical rejection probabilities are small when H_0 is true (model (4.1)). When H_0 is false, the powers of the τ_n and t tests are similar. Not surprisingly, the t test is somewhat more powerful than the τ_n test under model (4.2). The τ_n test is slightly more powerful under model (4.3).

5. AN EMPIRICAL EXAMPLE

This section presents an empirical example in which τ_n is used to test two hypotheses about the shape of an Engle curve. One hypothesis is that the curve is linear, and the other is that the curve is quadratic. The curve is given by (2.1). Y denotes the logarithm of the expenditure share of food consumed off the premises of the establishment where it was purchased, X denotes the logarithm of total expenditures, and W denotes annual income from wages and salaries. The data consist of 785 household-level observations from the 1996 U.S. Consumer Expenditure Survey. The bandwidth for estimating f_{XW} was selected by cross-validation. The kernel is the same as the one used in the Monte Carlo experiments. As in the experiments, the critical value of τ_n was estimated by setting $K_\varepsilon = 25$.

The τ_n test of the hypothesis that g is linear (quadratic) gives $\tau_n = 13.4$ (0.32) with a 0.05-level critical value of 3.07 (5.22). Thus, the test rejects the hypothesis that g is linear but not the hypothesis that g is quadratic. The hypotheses were also tested using the t test described in the section on Monte Carlo experiments. This test gives $t = 2.60$ for the hypothesis that g is linear ($\theta_2 = 0$ in (4.2)) and $t = 0.34$ for the hypothesis that g is quadratic ($\theta_3 = 0$ in (4.3)). The 0.05-level critical value is 1.96. Thus, the t test, like the τ_n test, rejects the hypothesis that g is linear but not the hypothesis that it is quadratic.

MATHEMATICAL APPENDIX: PROOFS OF THEOREMS

A.1 *Model* (2.1)

Define

$$S_{n1}(z) = n^{-1/2}\sum_{i=1}^{n} U_i f_{XW}(z, W_i),$$

$$S_{n2}(z) = n^{-1/2}\sum_{i=1}^{n} [g(X_i) - G(X_i, \theta_0)] f_{XW}(z, W_i),$$

$$S_{n3}(z) = n^{-1/2}\sum_{i=1}^{n} [G(X_i, \theta_0) - G(X_i, \hat{\theta}_n)] f_{XW}(z, W_i),$$

$$S_{n4}(z) = n^{-1/2}\sum_{i=1}^{n} U_i [\hat{f}_{XW}^{(-i)}(z, W_i) - f_{XW}(z, W_i)],$$

$$S_{n5}(z) = n^{-1/2}\sum_{i=1}^{n} [g(X_i) - G(X_i, \theta_0)][\hat{f}_{XW}^{(-i)}(z, W_i) - f_{XW}(z, W_i)],$$

and

$$S_{n6}(z) = n^{-1/2}\sum_{i=1}^{n}[G(X_i,\theta_0) - G(X_i,\hat{\theta}_n))][\hat{f}_{XW}^{(-i)}(z,W_i) - f_{XW}(z,W_i)] .$$

Then

$$S_n(z) = \sum_{j=1}^{6} S_{nj}(z) .$$

Lemma 1: As $n \to \infty$,

$$S_{n3}(z) = -\Gamma(z)' n^{1/2}(\hat{\theta}_n - \theta_0) + o_p(1)$$

$$= -\Gamma(z)' n^{-1/2}\sum_{i=1}^{n}\gamma(U_i, X_i, W_i, \theta_0) + o_p(1).$$

uniformly over $z \in [0,1]$.

Proof: A Taylor series expansion gives

$$S_{n3}(z) = -n^{-1}\sum_{i=1}^{n} G_\theta(X_i, \tilde{\theta}_n) f_{XW}(z, W_i) n^{1/2}(\hat{\theta}_n - \theta_0) ,$$

where $\tilde{\theta}_n$ is between $\hat{\theta}_n$ and θ_0. Application of Jennrich's (1969) uniform law of large numbers gives the first result of the lemma. The second result follows from the first by applying Assumption 3. Q.E.D.

Lemma 2: As $n \to \infty$, $|\partial \hat{f}_{XW}^{(-i)}(z,w)/\partial z - \partial f_{XW}(z,w)/\partial z| = o[(\log n)/(n^{1/2}h^2) + h]$ almost surely uniformly over $(z,w) \in [0,1]^2$.

Proof: This is a modified version of Theorem 2.2(2) of Bosq (1996) and is proved the same way as that theorem. Q.E.D.

Lemma 3: As $n \to \infty$, $S_{n4}(z) = o_p(1)$ uniformly over $z \in [0,1]$.

Proof: Let $I_1,...,I_m$ be a partition of $[0,1]$ into m intervals of length $1/m$. For each $j = 1,...,m$, choose a point $z_j \in I_j$. Define $\Delta f_{XW}^{(-i)}(x,w) = \hat{f}_{XW}^{(-i)}(x,w) - f_{XW}(x,w)$. Then for any $\varepsilon > 0$,

$$S_{n4}(z) = n^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{n} U_i I(z \in I_j)\Delta f_{XW}^{(-i)}(z, W_i)$$

$$= n^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{n} U_i I(z \in I_j)\Delta f_{XW}^{(-i)}(z_j, W_i)$$

$$+ n^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{n} U_i I(z \in I_j)[\Delta f_{XW}^{(-i)}(z, W_i) - \Delta f_{XW}^{(-i)}(z_j, W_i)]$$

$$\equiv S_{n41}(z) + S_{n42}(z).$$

A Taylor series expansion gives

$$S_{n42}(z) = n^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{n} U_i I(z \in I_j)[\partial \Delta f_{XW}^{(-i)}(\tilde{z}_j, W_i)/\partial z](z - z_j),$$

where $\tilde{z}_j$ is between z_j and z. Therefore, it follows from Lemma 2 that

$$| S_{n42}(z) | \le n^{-1/2} m^{-1}\sum_{j=1}^{m}\sum_{i=1}^{n} | U_i | I(z \in I_j) | \partial \Delta f_{XW}^{(-i)}(\tilde{z}_j, W_i)/\partial z |$$

$$\le n^{-1/2} m^{-1} o_p[(\log n)/(n^{1/2}h^2) + h]\sum_{j=1}^{m}\sum_{i=1}^{n} | U_i | I(z \in I_j)$$

$$= \quad O_p[(\log n)/(mh^2) + n^{1/2}h/m]$$

uniformly over $z \in [0,1]$. In addition, for any $\varepsilon > 0$,

$$\boldsymbol{P}\left[\sup_{z\in[0,1]} | S_{n41}(z) | > \varepsilon\right] = \boldsymbol{P}\left[\max_{1\le j\le m}\left| n^{-1/2}\sum_{i=1}^{n} U_i \Delta f_{XW}^{(-i)}(z_j, W_i)\right| > \varepsilon\right]$$

$$\le \sum_{j=1}^{m} \boldsymbol{P}\left[\left| n^{-1/2}\sum_{i=1}^{n} U_i \Delta f_{XW}^{(-i)}(z_j, W_i)\right| > \varepsilon\right].$$

But $\boldsymbol{E}[U_i \Delta f_{XW}^{(-i)}(z_j, W_i)] = 0$, and standard calculations for kernel estimators show that

$$Var\left[n^{-1/2}\sum_{i=1}^{n} U_i \Delta f_{XW}^{(-i)}(z, W_i)\right] = O[(nh^2)^{-1} + h^4]$$

for any $z \in [0,1]$. Therefore, it follows from Chebyshev's inequality that

$$\sum_{j=1}^{m} \boldsymbol{P}\left[\left| n^{-1/2} \sum_{i=1}^{n} U_i \Delta f_{XW}^{(-i)}(z_j, W_i) \right| > \varepsilon \right] = O[m/(nh^2\varepsilon^2) + mh^4/\varepsilon^2],$$

which implies that

$$\boldsymbol{P}\left[\sup_{z \in [0,1]} |S_{n41}(z)| > \varepsilon \right] = O[m/(nh^2\varepsilon^2) + mh^4/\varepsilon^2].$$

The lemma now follows by choosing m so that $n^{-1/2}m \to C_3$ as $n \to \infty$, where C_3 is a constant such that $0 < C_3 < \infty$. Q.E.D.

Lemma 4: As $n \to \infty$, $S_{n6}(z) = o_p(1)$ uniformly over $z \in [0,1]$.

Proof: A Taylor series expansion gives

$$S_{n6}(z) = n^{-1} \sum_{i=1}^{n} G_\theta(X_i, \tilde{\theta}_n)[\hat{f}_{XW}^{(-i)}(z, W_i) - f_{XW}(z, W_i)] n^{1/2}(\hat{\theta}_n - \theta_0),$$

where $\tilde{\theta}_n$ is between $\hat{\theta}_n$ and θ_0. The result follows from boundedness of G_θ, $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$, and $[\hat{f}_{XW}^{(-i)}(z, W_i) - f_{XW}(z, W_i)] = O[h^2 + (\log n)/(nh^2)^{1/2}]$ almost surely uniformly over $z \in [0,1]$. Q.E.D.

Lemma 5: Under H_0, $S_n(z) = B_n(z) + o_p(1)$ uniformly over $z \in [0,1]$.

Proof: Under H_0, $S_{n2}(z) = S_{n5}(z) = 0$ for all z. Now apply Lemmas 1, 2, and 4. Q.E.D.

Proof of Theorem 1:

Under H_0, $S_n(z) = B_n(z) + o_p(1)$ uniformly over $z \in [0,1]$ by Lemma 5. Under assumptions 1-4, Ω is a completely continuous operator, so its eigenvectors form a complete, orthonormal basis for $L_2[0,1]$. Therefore, $B_n(z)$ has the Fourier representation

$$B_n(z) = \sum_{j=1}^{\infty} b_j \psi_j(z),$$

where

$$b_j = \int_0^1 B_n(z) \psi_j(z) dz.$$

It follows that $\tau_n = \sum_{j=1}^{\infty} b_j^2 + o_p(1)$. Therefore, it suffices to find the asymptotic distribution of $\nu_n \equiv \sum_{j=1}^{\infty} b_j^2$.

To do this, observe that $\boldsymbol{E}(b_j^2)=\omega_j$ and $\boldsymbol{E}\nu_n \le C_\nu$ for some $C_\nu<\infty$. Therefore, for any $\varepsilon>0$ and $t\in(-\infty,\infty)$, there is a $K_\varepsilon<\infty$ such that $|t|\sum_{j=K_\varepsilon+1}^{\infty}\omega_j<\varepsilon$. Define $\nu_{nK}\equiv\sum_{j=1}^{K_\varepsilon} b_j^2$. For each j and k, $\boldsymbol{E}b_j=0$, and

$$\boldsymbol{E}b_j b_k=\boldsymbol{E}\int_0^1 dz_1\int_0^1 dz_2 B_n(z_1)B_n(z_2)\psi_j(z_1)\psi_k(z_2)$$

$$=\int_0^1 dz_1\int_0^1 dz_2 V(z_1,z_2)\psi_j(z_1)\psi_k(z_2)$$

$$=\int_0^1 \psi_j(z_1)(\Omega\psi_k)(z_1)dz_1$$

$$=\omega_j\delta_{jk},$$

where $\delta_{jk}=1$ if $j=k$ and 0 otherwise. It follows from the Lindeberg-Levy theorem that the b_j's are asymptotically independent $N(0,\omega_j)$ variates, and the random variables b_j^2/ω_j ($\omega_j\ne 0$) are independently chi-square distributed with one degree of freedom. Consequently,

$$\nu_{nK}\to^d \sum_{j=1}^{K_\varepsilon}\omega_j^2\chi_{1j}^2\equiv\eta_K .$$

Moreover,

(A1) $\quad |\boldsymbol{E}[\exp(\iota t\nu_{nK})-\exp(\iota t\eta_K)]|<\varepsilon$

for all sufficiently large n, where $\iota=\sqrt{-1}$.

Next, use the inequality $\left|e^{\iota t}-1\right|\le|t|$ to obtain

(A2) $\quad |\boldsymbol{E}[\exp(\iota t\nu_n)-\exp(\iota t\nu_{nK})|\le|t|\boldsymbol{E}|\nu_n-\nu_{nK}|$

$$=|t|\sum_{j=K_\varepsilon+1}^{\infty}\omega_j$$

$$<\varepsilon.$$

Define $\eta=\sum_{j=1}^{\infty}\omega_j\chi_{1j}^2$. Then

(A3) $|\boldsymbol{E}[\exp(\iota t\eta)-\exp(\iota t\eta_K)|\leq|t|\boldsymbol{E}|\eta-\eta_K|$

$$=|t|\boldsymbol{E}\sum_{j=K_\varepsilon+1}^{\infty}\omega_j\chi_{1j}^2$$

$$<\varepsilon.$$

Now combine (A1)-(A3) to obtain $|\boldsymbol{E}[\exp(\iota t\nu_n)-\exp(\iota t\eta)]|<\varepsilon$. Thus, the characteristic functions of ν_n and η can be made arbitrarily close by making n sufficiently large, which proves the theorem. Q.E.D.

Proof of Theorem 2: $|\hat{\omega}_j-\tilde{\omega}_j|=O\left(\left\|\hat{\Omega}-\tilde{\Omega}\right\|\right)$ by Theorem 5.1a of Bhatia, Davis, and McIntosh (1983). Moreover, standard calculations for kernel density estimators show that $\left\|\hat{\Omega}-\tilde{\Omega}\right\|=O[(\log n)/(nh^2)^{1/2}]$. Part (i) of the theorem follows by combining these two results. Part (ii) is an immediate consequence of part (i) and Theorem 1. Q.E.D.

Proof of Theorem 3: Let $\tilde{z}_\alpha$ denote the $1-\alpha$ quantile of the distribution of $\sum_{j=1}^{\infty}\tilde{\omega}_j\chi_{1j}^2$. It suffices to show that when H_1 holds, then under sampling from $Y=g(X)+U$,

$$\lim_{n\to\infty}\boldsymbol{P}(\tau_n>\tilde{z}_\alpha)=1.$$

This will be done by proving that

$$\operatorname*{plim}_{n\to\infty} n^{-1}\tau_n=\int_0^1[(Tq)(z)]^2\,dz>0.$$

To do this, observe that by Jennrich's (1969) uniform law of large numbers, $n^{-1/2}S_{n2}(z)=(Tq)(z)+o_p(1)$ uniformly over $z\in[0,1]$. Moreover, $S_{n5}(z)=o(h^{-1}\log n)=o(n^{1/6}\log n)$ a.s. uniformly over $z\in[0,1]$ because $\hat{f}_{XW}^{(-i)}(z,w)-f_{XW}(z,w)=o[(\log n)/(nh^2)^{1/2}]$ a.s. uniformly over $z\in[0,1]$. Combining these results with Lemma 5 yields

$$n^{-1/2}S_n(z)=n^{-1/2}B_n(z)+(Tq)(z)+o_p(1).$$

A further application of Jennrich's (1969) uniform law of large numbers shows that $n^{-1/2}S_n(z)\to^p(Tq)(z)$, so $n^{-1}\tau_n\to^p\int_0^1[(Tq)(z)]^2\,dz$. Q.E.D.

Proof of Theorem 4: Arguments like those leading to lemma 5 show that

$$S_n(z)=B_n(z)+S_{n2}(z)+S_{n5}(z)+o_p(1)$$

uniformly over $z\in[0,1]$. Moreover,

$$S_{n5}(z) = n^{-1}\sum_{i=1}^{n}\Delta(X_i)[\hat{f}_{XW}^{(-i)}(z,W_i) - f_{XW}(z,W_i)]$$

$$= O[(\log n)/(nh^2)^{1/2}]$$

almost surely uniformly over z. In addition

$$S_{n2}(z) = n^{-1}\sum_{i=1}^{n}\Delta(X_i) f_{XW}(z,W_i)$$

$$= \mu(z) + o(1)$$

almost surely uniformly over z. Therefore, $S_n(z) = B_n(z) + \mu(z) + o_p(1)$ uniformly over z. But

$$B_n(z) + \mu(z) = \sum_{j=1}^{\infty}\tilde{b}_j\psi_j(z),$$

where $\tilde{b}_j = b_j + \mu_j$ and b_j is defined as in the proof of Theorem 1. Moreover, the b_j's are asymptotically distributed as independent $N(\mu_j, \omega_j)$ variates. Now proceed as in the proof of Theorem 1. Q.E.D.

Proof of Theorem 5: Define $D_n(z) = S_{n3}(z) + S_{n6}(z) + \boldsymbol{E}[S_{n2}(z) + S_{n5}(z)]$ and $\tilde{S}_n(z) = S_n(z) - D_n(z)$. Then $\tau_n = \|\tilde{S}_n + D_n\|^2$. Use the inequality

(A5) $a^2 \geq 0.5b^2 - (b-a)^2$

with $a = S_n$ and $b = D_n$ to obtain

$$\boldsymbol{P}(\tau_n > z_\alpha) \geq \boldsymbol{P}\left(0.5\|D_n\|^2 - \|\tilde{S}_n\|^2 > z_\alpha\right).$$

For any finite $M > 0$,

$$\boldsymbol{P}\left(0.5\|D_n\|^2 - \|\tilde{S}_n\|^2 \leq z_\alpha\right) = \boldsymbol{P}\left(0.5\|D_n\|^2 \leq z_\alpha + \|\tilde{S}_n\|^2, \|\tilde{S}_n\|^2 \leq M\right)$$

$$+\boldsymbol{P}\left(0.5\|D_n\|^2 \leq z_\alpha + \|\tilde{S}_n\|^2, \|\tilde{S}_n\|^2 > M\right)$$

$$\leq \boldsymbol{P}\left(0.5\|D_n\|^2 \leq z_\alpha + M\right) + \boldsymbol{P}\left(\|\tilde{S}_n\|^2 > M\right).$$

$\|\tilde{S}_n\|$ is bounded in probability uniformly over $g \in \mathcal{F}_{nc}$. Therefore, for each $\varepsilon > 0$ there is $M_\varepsilon < \infty$ such that for all $M > M_\varepsilon$

$$P\left(0.5\|D_n\|^2 - \|\tilde{S}_n\|^2 \le z_\alpha\right) \le P\left(.5\|D_n\|^2 \le z_\alpha + M\right) + \varepsilon .$$

Equivalently,

$$P\left(0.5\|D_n\|^2 - \|\tilde{S}_n\|^2 > z_\alpha\right) \ge P\left(.5\|D_n\|^2 > z_\alpha + M\right) - \varepsilon$$

and

(A6) $$P(\tau_n > z_\alpha) \ge P\left(.5\|D_n\|^2 > z_\alpha + M\right) - \varepsilon .$$

Now

$$S_{n2}(z) + S_{n5}(z)] = n^{-1/2}\sum_{i=1}^{n}[g(X_i) - G(X_i,\theta_g)]\hat{f}_{XW}^{(-i)}(z,W_i) .$$

Therefore,

$$E[S_{n2}(z) + S_{n5}(z)] = n^{-1/2}E\sum_{i=1}^{n}[g(X_i) - G(X_i,\theta_g)][f_{XW}(z,W_i) + h^2 R_n(z)] ,$$

where $R_n(z)$ is nonstochastic, does not depend on g, and is bounded uniformly over $z \in [0,1]$. It follows that

$$E[S_{n2}(z) + S_{n5}(z)] = n^{1/2}(Tq)(z) + O\left[n^{1/2}h^2\|q\|\right]$$

and

$$E[S_{n2}(z) + S_{n5}(z)] \ge 0.5n^{1/2}(Tq)(z)$$

uniformly over $g \in \mathcal{F}_{nc}$ for all sufficiently large n.

Now

$$\left|S_{n3}(z) + S_{n6}(z)\right| \le \sup_{x\in[0,1], g\in\mathcal{F}_{nc}} n^{1/2}\left|G(x,\hat{\theta}_n) - G(x,\theta_g)\right| n^{-1}\sum_{i=1}^{n}\hat{f}_{XW}^{(-i)}(z,W_i) .$$

Therefore, it follows from the definition $\mathcal{F}_{nc}$ and uniform convergence of $\hat{f}_{XW}^{(-i)}$ to f_{XW} that $\|S_{n3} + S_{n5}\| = O_p(1)$ uniformly over $g \in \mathcal{F}_{nc}$. A further application of (A5) with $a = D_n(z)$ and $b = E[S_{n2}(z) + S_{n5}(z)]$ gives

(A7) $$\|D_n\|^2 \ge .125n\|Tq\|^2 + O_p(1)$$

uniformly over $g \in \mathcal{F}_{nc}$ as $n \to \infty$. The theorem follows by substituting (A7) into (A6) and choosing C to be sufficiently large. Q.E.D.

A.2 *Model* (3.1)

Proofs of Theorems 6-10

The proofs are identical to the proofs of Theorems 1-5 after replacing quantities for model (2.1) with the analogous quantities for model (3.1). Q.E.D.

FOOTNOTES

[1] Tests of a parametric model of a conditional mean or quantile function against a nonparametric alternative include Aït-Sahalia, *et al*. (1994), Bierens (1982, 1990), Bierens and Ginther (2000), Bierens and Ploberger (1997), de Jong (1996), Eubank and Spiegelman (1990), Fan (1996), Fan and Huang (2001), Fan and Li (1996), Gozalo (1993), Guerre and Lavergne (2002), Härdle and Mammen (1993), Hart (1997), Hong and White (1995), Horowitz and Spokoiny (2001, 2002), Stute (1997), Li and Wang (1998), Whang and Andrews (1993), Wooldridge (1992), Yatchew (1992), and Zheng (1996, 1998).

[2] At the cost of additional analytic complexity, it may be possible to let $\varepsilon \to 0$ and $K_\varepsilon \to \infty$ as $n \to \infty$, thereby obtaining a consistent estimator of the asymptotic critical value of τ_n. However, doing this would likely provide little insight into the accuracy of the estimator or the choice of K_ε in applications. This is because the difference between the distributions of $\tilde{\tau}$ and $\tilde{\tau}_\varepsilon$ and, therefore, the approximation error are complicated functions of the multiplicities and spacings of the $\tilde{\omega}_j$'s. Letting $\varepsilon \to 0$ and $K_\varepsilon \to \infty$ has no practical consequences because $\varepsilon > 0$ and $K_\varepsilon < \infty$ with any finite sample.

REFERENCES

Aït-Sahalia, Y., P.J. Bickel, and T.M. Stoker (2001). Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities, *Journal of Econometrics*, 105, 363-412.

Andrews, D.W.K. (1997). A Conditional Kolmogorov Test, *Econometrica*, 65, 1097-1128.

Bhatia, R., C. Davis, and A. McIntosh (1983). Perturbation of Spectral Subspaces and Solution of Linear Operator Equations, *Linear Algebra and Its Applications*, 52/53, 45-67.

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. and D.K. Ginther (2000). Integrated Conditional Moment Testing of Quantile Regression Models, working paper, Department of Economics, Pennsylvania State University.

Bierens, H.J. and W. Ploberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica*, 65, 1129-1151.

Blundell, R., X. Chen and D. Kristensen (2003). Semi-Nonparametric IV Estimation of Shape Invariant Engle Curves, working paper CWP 15/03, Centre for Microdata Methods and Practice, University College London.

Bosq, D. (1996). Nonparametric Statistics for Stochastic Processes, New York: Springer.

Darolles, S., J.-P. Florens, and E. Renault (2002). Nonparametric Instrumental Regression, working paper, GREMAQ, University of Social Science, Toulouse.

de Jong, R.M. (1996). On the Bierens Test under Data Dependence, *Journal of Econometrics*, 72, 1-32.

Eubank, R.L. and C.H. 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. (1996). Test of Significance Based on Wavelet Thresholding and Neyman's Truncation, *Journal of the American Statistical Association*, 91, 674-688.

Fan, J. and L.-S. Huang (2001). Goodness of Fit Tests for Parametric Regression Models, *Journal of the American Statistical Association*, 96, 640-652.

Fan, Y. and Q. Li (1996). Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, *Econometrica*, 64, 865-890.

Gozalo, P.L. (1993). A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models, *Econometric Theory*, 9, 451-477.

Guerre, E. and P. Lavergne (2002). Optimal Minimax Rates for Nonparametric Specification Testing in Regression Models, *Econometric Theory*, 18, 1139-1171.

Hall, P. and J.L. Horowitz (2003). Nonparametric Methods for Inference in the Presence of Instrumental Variables, working paper, Department of Economics, Northwestern University.

Härdle, W. and E. Mammen (1993). Comparing Nonparametric Versus Parametric Regression Fits, *Annals of Statistics*, 21, 1926-1947.

Hart, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. New York: Springer-Verlag.

Hong, Y. and H. White (1996). Consistent Specification Testing via Nonparametric Series Regressions, *Econometrica*, 63, 1133-1160.

Horowitz, J.L. and V.G. Spokoiny (2001). An Adaptive, Rate-Optimal Test of a Parametric Mean Regression Model against a Nonparametric Alternative, *Econometrica*, 69, 599-631.

Horowitz, J.L. and V.G.Spokoiny (2002). An Adaptive, Rate-Optimal Test of Linearity for Median Regression Models, *Journal of the American Statistical Association*, 97, 822-835.

Jennrich, R.I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators, *Annals of Mathematical Statistics*, 40, 633-643.

Li, Q. and S. Wang (1998). A Simple Consistent Bootstrap Test for a Parametric Regression Function, *Journal of Econometrics*, 87, 145-165.

Newey, W.K. and J.L. Powell (2003). Instrumental Variable Estimation of Nonparametric Models, *Econometrica*, 71, 1565-1578.

Newey, W.K., J.L. Powell, and F. Vella (1999). Nonparametric Estimation of Triangular Simultaneous Equations Models, *Econometrica*, 67, 565-603.

Stute, W. (1997). Nonparametric Model Checks for Regression, *Annals of Statistics*, 25, 613-641.

Whang, Y.-J. and D.W.K. Andrews (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.

Yatchew, A.J. (1992). Nonparametric Regression Tests Based on Least Squares, *Econometric Theory*, 8, 435-451.

Zheng, J.X. (1996). A Consistent Test of Functional Form via Nonparametric Estimation Techniques, *Journal of Econometrics*, 75, 263-289.

Zheng, J.X. (1998). A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions, *Econometric Theory*, 14, 123-138.

Table 1: Results of Monte Carlo Experiments

Model	ρ	η	Empirical Probability that H_0 Is Rejected Using τ_n	t test
			H_0 is true	
(11)	0.8	0.1	0.051	0.052
	0.8	0.5	0.030	0.034
	0.7	0.1	0.049	0.052
			H_0 is false	
(12)	0.8	0.1	0.658	0.714
	0.8	0.5	0.721	0.827
	0.7	0.1	0.421	0.444
(13)	0.8	0.1	0.684	0.671
	0.8	0.5	0.663	0.580
	0.7	0.1	0.424	0.412