Testing a Parametric Model against a Nonparametric Alternative with Identification through Instrumental Variables

Joel Horowitz

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP14/04

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

September 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 a 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 and has asymptotic power advantages over existing tests. Moreover, it has power arbitrarily close to 1 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 Donald Andrews and Richard Blundell for helpful discussions and comments. Research supported in part by NSF Grants SES 991092 and SES 0352675.

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

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

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 and has asymptotic power advantages over existing tests. In large samples its power is arbitrarily close to 1 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). Methods for testing (1.2) against a nonparametric alternative have been developed by Donald, Imbens, and Newey (2003) and Tripathi and Kitamura (2003). In addition, the test of a conditional mean function developed by Bierens (1990) and Bierens and Ploberger (1997) can be modified to provide a test of (1.2). Horowitz and Spokoiny (2001,2002) provide extensive references to other tests for conditional mean and quantile functions. The test presented here has asymptotic power advantages over existing tests that permit X to have endogenous components. In addition, among existing tests of (1.2) against a nonparametric alternative, only the test presented here is uniformly consistent at a known rate over a known set of alternative hypotheses. Uniform consistency is important because it provides some assurance that there are not alternatives against which a test has low power even with large samples. If a test is not uniformly consistent over a specified set, then that set contains alternatives against which the test has low power.

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 is not affected by the ill-posed inverse problem of nonparametric instrumental variables estimation. Consequently, the test's "precision" exceeds that of any nonparametric estimator of g. The rate of convergence of a nonparametric estimator of g is always slower than $O_p(n^{-1/2})$ and, depending on the details of the 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 here 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 rate of estimation. See Guerre and Lavergne (2002) and Horowitz and Spokoiny (2001, 2002).

Section 2 describes the test statistic and its properties. Section 3 presents the results of a Monte Carlo investigation of the finite-sample performance of the test, and Section 4 presents an illustrative application to real data. The proofs of theorems are in the appendix.

2. THE TEST STATISTIC AND ITS PROPERTIES

Rewrite (1.1) as

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

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 included in (2.1). X and Z, respectively, are endogenous and exogenous explanatory variables. W is an instrument for X. The assumption that $\text{supp}(X,Z,W) \subset [0,1]^{2p+r}$ can always be satisfied by carrying out a monotone transformation of (X,Z,W). The inferential problem is to test the null hypothesis, H_0, that

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

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 (2.2) 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).

2.1 *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

$$(2.3) \qquad \tilde{S}(x,z) \equiv T_z[g(\cdot,\cdot) - G(\cdot,\cdot,\theta)](x,z) = 0$$

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 (2.3) holds. 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 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 $L_2[0,1]^r$ if $r > 0$. That is, the operator L defined by

$$L\nu(z) = \int \ell(\zeta, z) \nu(\zeta) d\zeta$$

is nonsingular. Let L be the identity operator if $r = 0$. Define the operator T on $L_2[0,1]^{p+r}$ by $T\nu(x,z) = LT_z \nu(x,z)$. Then T is non-singular. H_0 is equivalent to

$$(2.4) \qquad S(x,z) \equiv T[g(\cdot,\cdot) - G(\cdot,\cdot,\theta)](x,z) = 0$$

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 (2.4) holds. The test statistic is based on a sample analog of

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

To form the analog, observe that under H_0,

$$T[g - G(\cdot,\cdot,\theta)](x,z) = \boldsymbol{E}\{[Y - G(X,Z,\theta)]f_{XW}(x,z,W)\ell(Z,z\}.$$

Therefore, it suffices to form a sample analog of $\boldsymbol{E}\{[Y - G(X,Z,\theta)]f_{XW}(x,z,W)\ell(Z,z\}$. To do this, 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 $S(x,z)$ is

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

The test statistic is

$$\tau_n = \int S_n^2(x,z)\,dxdz \qquad (2.5)$$

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. 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). Let $D_j f_{XZW}$ denote any j'th partial or mixed partial derivative of f_{XZW}. Let $D_0 f_{XZW}(x,z,w) = f_{XZW}(x,z,w)$.

1. (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 $z \in [0,1]^r$.

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

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 $g(x,z) = G(x,z,\theta_0)$, $\theta_0 \in \text{int}(\Theta)$, and

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

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.

4. (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 $\theta \in \Theta$.

5. (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.

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 $G_\theta(x,z,\theta) = \partial G(x,z,\theta)/\partial\theta$, $\Gamma(x,z) = \boldsymbol{E}[G_\theta(X,Z,\theta_0) f_{XZW}(x,Z,W)\ell(Z,z)]$,

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

and $V(x_1,z_1;x_2,z_2) = \boldsymbol{E}[B_n(x_1,z_1)B_n(x_2,z_2)]$. Define the operator Ω on $L_2[0,1]^{q+r}$ by

$$(2.7) \qquad (\Omega v)(x,z) = \int_0^1 V(x,z;\xi,\zeta)v(\xi,\zeta)d\xi d\zeta .$$

Let $\{\omega_j : j = 1,2,...\}$ denote the eigenvalues of Ω sorted so that $\omega_1 \ge \omega_2 \ge ... \ge 0$. Let $\{\chi^2_{1j} : j = 1,2,...\}$ 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: If H_0 is true and assumptions 1-5 hold, then

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

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 replaces 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 $n \to \infty$. The 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 $\boldsymbol{E}[Y - G(X,\theta_0) \mid Z,W] = 0$ 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

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

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 $\hat{\theta}_n$. This model coincides with (2.1) when H_0 is true. Moreover, H_0 holds for the pseudo-true model in the sense that $\boldsymbol{E}[\tilde{Y} - G(X,Z,\theta_0) \mid Z,W] = 0$, regardless of whether H_0 holds for (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 (2.8). Order the $\tilde{\omega}_j$'s such that $\tilde{\omega}_1 \geq \tilde{\omega}_2 \geq ... \geq 0$. Then under sampling from (2.8), 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

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

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 H_0 is rejected. Similarly, use of the approximation creates an arbitrarily small change in the power of the τ_n test when H_0 is false. However, 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 (2.8). 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.

The remainder of this section describes how to obtain the estimated eigenvalues $\{\hat{\omega}_j\}$. 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

$$\text{(2.9)} \qquad \hat{\theta}_n = \arg\min_{\theta\in\Theta} \left\{ \sum_{i=1}^{n} [Y_i - G(X_i, Z_i, \theta)] \tilde{W}_i' \right\} A_n \left\{ \sum_{i=1}^{n} [Y_i - G(X_i, Z_i, \theta)] \tilde{W}_i \right\},$$

where $\{A_n\}$ is a sequence of possibly stochastic $c_\theta \times c_\theta$ weight matrices converging in probability to a finite, non-stochastic 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 $\tilde{\gamma} = (D'AD)^{-1} D'A$. Then standard calculations for GMM estimators show that

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

Therefore,

$$\text{(2.10)} \qquad V(x_1, z_1; x_2, z_2) = \boldsymbol{E} \left\{ n^{-1} \sum_{i=1}^{n} [f_{XZW}(x_1, Z_i, W_i) \ell(Z_i, z_1) - \Gamma(x_1, z_1)' \tilde{\gamma} \tilde{W}_i] \tilde{U}_i^2 \right.$$

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

A consistent estimator of V can be obtained by replacing unknown quantities on the right-hand side of (2.10) with estimators. To this end, define $\hat{D} = n^{-1}\sum_{i=1}^{n}\tilde{W}_i G_\theta(X,Z,\hat{\theta}_n)'$, $\hat{\gamma} = (\hat{D}'A_n\hat{D})^{-1}\hat{D}'A_n$, 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 $\hat{U}_i = Y_i - G(X_i,Z_i,\hat{\theta}_n) - \hat{q}^{(-i)}(Z_i,W_i)$, where $\hat{q}^{(-i)}(z,w)$ is the leave-observation i-out kernel regression estimator of $Y - G(X,Z,\hat{\theta}_n)$ on (Z,W). Then $V(x_1,z_1;x_2,z_2)$ is estimated consistently by

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

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

Let $\hat{\Omega}$ be the integral operator whose kernel is $\hat{V}(x_1,z_1;x_2,z_2)$. The $\hat{\omega}_j$'s are the eigenvalues of $\hat{\Omega}$.

Theorem 2: Let assumptions 1-5 hold. Then as $n \to \infty$, (i) $\sup_{1\le j\le K_\varepsilon}|\hat{\omega}_j - \tilde{\omega}_j| = O[(\log n)/(nh^{2p+r})^{1/2}]$ 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 $\hat{F}(x,z)$ denote the $n\times 1$ vector whose i'th component is $\hat{f}_{XW}(x,Z_i,W_i)\ell(Z_i,z)$, $\hat{G}_\theta$ denote the $n\times d$ matrix whose (i,j) element is $G_\theta(X_i,Z_i,\hat{\theta}_n)$, 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 $M = I_n - n^{-1}\hat{G}_\theta\hat{\gamma}\tilde{W}'$, where I_n is the $n\times n$ identity matrix. Then

$$\hat{V}(z_1,z_2) = n^{-1}\hat{F}(x_1,z_1)'M\Upsilon M'\hat{F}(x_2,z_2).$$

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

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

where

$$\hat{d}_{jk} = \int_0^1 dx \int_0^1 dz_1 \int_0^1 dz_2 \int_0^1 dw \hat{f}_{XZW}(x, z_1, w)\ell(z_1, z_2)\phi_j(x, z_1)\phi_k(z_2 w) .$$

Approximate $\hat{f}_{XZW}(x, z, W)\ell(Z, z)$ by the finite sum

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

for some integer $L < \infty$. Since $\hat{f}_{XZW}\ell$ is a known function, L can be chosen to make Π approximate $\hat{f}_{XZW}\ell$ with any desired accuracy. Let $\phi(x, z)$ denote the $L \times 1$ vector whose j'th component is $\phi_j(x, z)$. Let Φ be the $L \times n$ matrix whose (j, k) component is $\phi_j(Z_k, W_k)$. Let D be the $L \times L$ matrix $\{d_{jk}\}$. Then $\hat{V}(x_1, z_1; x_2, z_2)$ is approximated by

$$\hat{\hat{V}}(x_1, z_1; x_2, z_2) = n^{-1}\phi(x_1, z_1)' D\Phi M \Upsilon M' \Phi' D' \phi(x_2, z_2) .$$

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 $g(x, z) = G(x, z, \theta)$ for almost every (x, z). Let θ_0 denote the probability limit of $\hat{\theta}_n$. Define $q(x, z) = g(x, z) - G(x, z, \theta_0)$. Let $\tilde{z}_\alpha$ denote the $1 - \alpha$ quantile of the distribution of τ_n under sampling from the pseudo-true model (2.8). 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: Let assumptions 1-5 hold. Suppose that H_0 is false and that $\int_0^1 [(Tq)(x, z)]^2 dxdz > 0$. Then for any α such that $0 < \alpha < 1$,

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

and

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

Because T is nonsingular, the τ_n test is consistent against any alternative that differs from $G(x, z, \theta_0)$ on a set of (x, z) values whose Lebesgue measure exceeds zero. It is shown in the proof of Theorem 3 that when H_0 is false, τ_n increases at the rate n as n increases. This is faster than the rate found by Donald, Imbens, and Newey (2003) for a test of H_0 against a nonparametric alternative based on the GMM test of overidentifying restrictions.

2.6 *Asymptotic Distribution under Local Alternatives*

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

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

where Δ is a bounded function on $[0,1]^{p+r}$ and $\theta_0 \in \text{int}(\Theta)$. The following additional notation is used. Let $\hat{\theta}_n$ be the GMM estimator (2.9). Let $\{\psi_j : j = 1,2,...\}$ denote the orthonormal eigenvectors of Ω. Define $\mu(x,z) = T\{\Delta - \boldsymbol{E}[W\Delta(X,Z)]'\tilde{\gamma}'G_\theta\}(x,z)$ and

$$\mu_j = \int_0^1 \mu(x,z)\psi_j(x,z)dxdz .$$

Let $\{\chi^2_{1j}(\mu_j^2/\omega_j) : j = 1,2,...\}$ denote independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $\{\mu_j^2/\omega_j\}$. Let $\hat{\theta}_n$ be the GMM estimator (2.9). The following theorem states the result.

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

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

where the ω_j 's are the eigenvalues of the operator Ω defined in (2.7).

Let z_α denote the $1-\alpha$ quantile of the distribution of $\sum_{j=1}^{\infty} \omega_j \chi^2_{1j}(\mu_j^2/\omega_j)$. 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 $\varepsilon > 0$,

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

It also follows from Theorem 4 that the τ_n test has power against local alternatives whose distance from the null-hypothesis model is $O(n^{-1/2})$. In contrast, the test of Tripathi and Kitamura (2003) has power only against local alternatives whose distance from the null-hypothesis model decreases more slowly than $n^{-1/2}$. If $\mu(x,z) = 0$ for all (x,z), then there is a non-stochastic sequence $\{\theta_n\}$ such that $G(x,z,\theta_n) = G(x,z,\theta_0) + n^{-1/2}\Delta(x,z) + o(n^{-1/2})$. Therefore, the distance between the null and alternative hypotheses is $o(n^{-1/2})$.

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 $g(x,z) = G(x,z,\theta)$ for some $\theta \in \Theta$ and a given function G. Let $\tilde{\Theta}$ be a compact subset of $\text{int}(\Theta)$. Define $q_g(x,z) = g(x,z) - G(x,z,\theta_g)$. Let h denote the bandwidth in $f_{XZW}^{(-i)}$. For each $n = 1,2,...$ and $C > 0$ define $\mathcal{F}_{nc}$ 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) $\theta_g \in \tilde{\Theta}$; (iii) (2.6) holds uniformly over $g \in \mathcal{F}_{nc}$; (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 (ii) insures the existence of the critical value defined in Section 2.4. The requirement $\theta_g \in \tilde{\Theta}$ is not restrictive in applications because Θ and $\tilde{\Theta}$ can usually be made large enough to include any reasonable θ_g. 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 $p = 1$, $r = 0$, and $\{\lambda_j, \phi_j : j = 1,2,...\}$ denote the eigenvalues and eigenvectors of T ordered so that $\lambda_1 \ge \lambda_2 \ge ... > 0$. Let, $G(x,\theta) = \theta\phi_1(x)$, $g(x) = \phi_1(x) + \phi_n(x)$, 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 H_0 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. Assume that $\hat{\theta}_n$ satisfies (2.9). Then given any $\delta > 0$, α such that $0 < \alpha < 1$, and any sufficiently large but finite constant C,

$$\lim_{n\to\infty} \inf_{\mathcal{F}_{nc}} \boldsymbol{P}(\tau_n > z_\alpha) \ge 1 - \delta \,. \tag{2.12}$$

and

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

2.8 *Alternative Weights*

This section compares τ_n with a generalization of the test of Bierens (1990) and Bierens and Ploberger (1997). To minimize the complexity of the discussion, assume that $p=1$ and $r=0$, so Z is not in the model. Let H be a bounded, real-valued function on $[0,1]^2$ such that

$$\left\| \int_0^1 H(x,w)s(w)dw \right\|^2 = 0$$

only if $s(w)=0$ for almost every $w \in [0,1]$. Then a test of H_0 can be based on the statistic

$$\tau_{nH} = \int_0^1 S_{nH}^2(x)dx ,$$

where

$$S_{nH}(x) = n^{-1/2} \sum_{i=1}^{n} [Y_i - G(X_i, \hat{\theta}_n)] H(x, W_i) .$$

If $H(x,w) = \tilde{H}(xw)$ for a suitably chosen function $\tilde{H}$, then τ_{nH} is a modification of the statistic of Bierens (1990) and Bierens and Ploberger (1997) for testing the hypothesis that a conditional mean function belongs to a specified, finite-dimensional parametric family. In this section, it is shown that the power of the τ_{nH} test can be low relative to that of the τ_n test. Specifically, there are combinations of density functions of (X,W), f_{XW}, and local alternative models (2.11) such that an α-level τ_{nH} test based on a fixed H has asymptotic local power arbitrarily close to α, whereas the α-level τ_n test has asymptotic local power that is bounded away from α. The opposite situation cannot occur under the assumptions of this paper. That is, it is not possible for the asymptotic power of the α-level τ_n test to approach α while the power of the α-level τ_{nH} test remains bounded away from α.

The conclusion that the power of τ_{nH} can be low relative to that of τ_n is reached by constructing an example in which the α-level τ_n test has asymptotic power that is bounded away from α but the τ_{nH} test has asymptotic power that is arbitrarily close to α. To minimize the complexity of the example, assume that θ is known and does not have to be estimated. Define

$$\bar{B}_n(x) = n^{-1/2} \sum_{i=1}^{n} U_i f_{XW}(x, W_i) ,$$

$$\bar{B}_{nH}(x) = n^{-1/2} \sum_{i=1}^{n} U_i H(x, W_i) ,$$

$\bar{R}(x_1,x_2)=\boldsymbol{E}[\bar{B}_n(x_1)\bar{B}_n(x_2)]$, and $\bar{R}_H(x_1,x_2)=\boldsymbol{E}[\bar{B}_{nH}(x_1)\bar{B}_{nH}(x_2)]$. Also, define the operators $\bar{\Omega}$ and $\bar{\Omega}_H$ on $L_2[0,1]$ by

$$(\bar{\Omega}\psi)(x)=\int_0^1 \bar{R}(x,\xi)\psi(\xi)d\xi$$

and

$$(\bar{\Omega}_H\psi)(x)=\int_0^1 \bar{R}_H(x,\xi)\psi(\xi)d\xi .$$

Let $\{\bar{\varpi}_j,\bar{\psi}_j : j=1,2,...\}$ and $\{\bar{\varpi}_{jH},\bar{\psi}_{jH} : j=1,2,...\}$ denote the eigenvalues and eigenvectors of $\bar{\Omega}$ and $\bar{\Omega}_H$, respectively, with the eigenvalues sorted in decreasing order. For Δ defined as in (2.11), define $\bar{\mu}(x)=(T\Delta)(x)$,

$$\bar{\mu}_H(x)=\int_0^1\int_0^1 \Delta(\xi)H(x,w)f_{XW}(\xi,w)dxdw ,$$

$$\bar{\mu}_j=\int_0^1 \bar{\mu}(x)\bar{\psi}_j(x)dx ,$$

and

$$\bar{\mu}_{jH}=\int_0^1 \bar{\mu}_H(x)\bar{\psi}_{jH}(x)dx .$$

Then arguments like those used to prove Theorem 4 show that under the sequence of local alternatives (2.11) with a known θ,

$$\tau_n \to^d \sum_{j-1}^{\infty}\bar{\varpi}_j\chi_{1j}^2(\bar{\mu}_j^2/\bar{\varpi}_j)$$

and

$$\tau_{nH} \to^d \sum_{j-1}^{\infty}\bar{\varpi}_{jH}\chi_{1j}^2(\bar{\mu}_{jH}^2/\bar{\varpi}_{jH})$$

as $n\to\infty$. Therefore, to establish the first conclusion of this section, it suffices to show that for a fixed function H, f_{XW} and Δ can be chosen so that $\|\bar{\mu}\|^2/\sum_{j=1}^{\infty}\bar{\varpi}_j$ is bounded away from 0 and $\|\bar{\mu}_H\|^2/\sum_{j=1}^{\infty}\bar{\varpi}_{jH}$ is arbitrarily close to 0.

To this end, let $\phi_1(x)=1$ and $\phi_{j+1}(x)=2^{-1/2}\cos(j\pi x)$ for $j\ge 1$. Let $m>1$ be a finite integer. Define

$$\lambda_j=\begin{cases}1 \text{ if } j=1 \text{ or } m\\ e^{-2j} \quad \text{otherwise.}\end{cases}$$

Let

$$f_{XW}(x,w) = 1 + \sum_{j=1}^{\infty} \lambda_{j+1}^{1/2} \phi_{j+1}(x) \phi_{j+1}(w) .$$

Let $\boldsymbol{E}(U^2 | W = w) = 1$ for all $w \in [0,1]$. Then $\bar{R}(x_1, x_2) = t(x_1, x_2)$, $\bar{\omega}_j = \lambda_j$, and $\sum_{j=1}^{\infty} \bar{\omega}_j$ is non-zero and finite. Set $\Delta(x) = D\phi_m(x)$ for some finite $D > 0$. Then $\|\bar{\mu}\|^2 = D^2 \lambda_m^2 = D^2$. It suffices to show that m can be chosen so that $\|\bar{\mu}_H\|^2$ is arbitrarily close to 0. To do this, observe that $H(z,w)$ has the Fourier representation

$$H(x,w) = \sum_{j,k=1}^{\infty} h_{jk} \phi_j(x) \phi_k(w) ,$$

where $\{h_{jk} : j,k = 1,2,...\}$ are constants. Moreover, $\|\bar{\mu}_H\|^2 = D^2 \sum_{j=1}^{\infty} h_{jm}^2$. Since H is bounded, m can be chosen so that $\sum_{j=1}^{\infty} h_{jm}^2 < \varepsilon / D^2$ for any $\varepsilon > 0$. With this m, $\|\bar{\mu}_H\|^2 < \varepsilon$, which establishes the first conclusion.

The opposite situation (a sequence of local alternatives for which $\|\bar{\mu}\|^2$ approaches 0 while is $\|\bar{\mu}_H\|^2$ remains bounded away from 0) cannot occur. To show this, assume without loss of generality that the marginal distributions of X and W are $U[0,1]$, $\boldsymbol{E}(U^2 | W = w) = 1$ for all $w \in [0,1]$, and $\sum_{j=1}^{\infty} \bar{\omega}_{jH} = 1$. Also, assume that $\|\Delta\|^2 < C_\Delta$ for some constant $C_\Delta < \infty$. Then,

$$\int_0^1 \int_0^1 H(x,w)^2 \, dx dw = \sum_{j=1}^{\infty} \bar{\omega}_{jH} .$$

It follows from the Cauchy-Schwartz inequalty that

$$\begin{aligned}
\|\bar{\mu}_H\|^2 &\leq \left[\int_0^1 \int_0^1 H(x,w)^2 \, dz dw \right] \int_0^1 \left[\int_0 f_{XW}(x,w) \Delta(x) dx \right]^2 dw \\
&= \int_0^1 \left[\int_0 f_{XW}(x,w) \Delta(x) dx \right]^2 dw \\
&\leq \|\Delta\|^2 \|T\Delta\|^2 \\
&\leq C_\Delta \|\bar{\mu}\|^2 .
\end{aligned}$$

Therefore, $\|\bar{\mu}\|^2$ can approach 0 only if $\|\bar{\mu}_H\|^2$ also approaches 0.

3. 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 null hypotheses, H_0, that

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

and

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

The alternative hypotheses are (4.2) if (4.1) is H_0 and

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

if either (4.1) or (4.2) is H_0.

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 two other tests. One is an asymptotic t test of the hypothesis $\theta_2 = 0$ if (4.1) is H_0 and of $\theta_3 = 0$ if (4.2) is H_0. The t test is an example of an *ad hoc* test that might be used in applied research. The other test is the modified test of Bierens (1990) and Bierens and Ploberger (1997) that is described in Section 2.8. The weight function is $H(x,w) = \exp(xw)$. The critical value was computed using the methods described in Section 2.4.

In all experiments, $\theta_0 = 0$ and $\theta_1 = 0.5$. When (4.2) is the correct model, $\theta_2 = -0.5$. When (4.3) is the correct model, $\theta_2 = -1$, $\theta_3 = 1$ if (4.1) is H_0, and $\theta_3 = 4$ if (4.2) is H_0. 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 are not sensitive to the choice of K_ε, and the estimated eigenvalues $\hat{\omega}_j$ are very close to 0 when $j > 25$. The sample size is $n = 500$, and the nominal level is 0.05. There are 1000 Monte Carlo replications in each experiment. Computation of the critical value took approximately 4 seconds on a 900 MHz PC.

The results are shown in Table 1. The differences between the nominal and empirical rejection probabilities are small when H_0 is true. When H_0 is false, the powers of the τ_n and t tests are similar. Not surprisingly, the t tests of (4.1) against (4.2) and (4.2) against (4.3) are somewhat more powerful than the τ_n tests. The τ_n test is slightly more powerful for testing (4.1) against (4.3). The Bierens-type test is much less powerful than the τ_n and t tests, especially for testing (4.2) against (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 is that the curve is linear. The other is that it is quadratic. The curve is given by (2.1) with $r = 0$, so Z is not in the model. Y denotes the logarithm of the expenditure share of food consumed off the premises 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, $K_\varepsilon = 25$.

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 that g is quadratic. The hypotheses were also tested using the t test described in Section 4. 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 also rejects the hypothesis that g is linear but not the hypothesis that it is quadratic.

MATHEMATICAL APPENDIX: PROOFS OF THEOREMS

To minimize the complexity of the presentation, it is assumed here that $p = 1$ and $r = 0$. The proofs for $p > 1$ and/or $r > 0$ are identical after replacing quantities for $p = 1, r = 0$ with the analogous quantities for the more general case. Let f_{XW} denote the density function of (X, W).

Define

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

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

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

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

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

and

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

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

<u>Lemma 1</u>: As $n \to \infty$,

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

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

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

<u>Proof</u>: A Taylor series expansion gives

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

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.

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

<u>Proof</u>: 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.

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

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 $x_j \in I_j$. Define $\Delta f_{XW}^{(-i)}(x,w) = \hat{f}_{XW}^{(-i)}(x,w) - f_{XW}(x,w)$. Then for any $\varepsilon > 0$,

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

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

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

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

A Taylor series expansion gives

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

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

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

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

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

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

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

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

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

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

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

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

which implies that

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

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 $n\to\infty$, $S_{n6}(x)=o_p(1)$ uniformly over $x\in[0,1]$.

Proof: A Taylor series expansion gives

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

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 $[\hat{f}_{XW}^{(-i)}(x,W_i)-f_{XW}(x,W_i)]=O[h^2+(\log n)/(nh^2)^{1/2}]$ almost surely uniformly over $x\in[0,1]$. Q.E.D.

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

Proof: Under H_0, $S_{n2}(x)=S_{n5}(x)=0$ for all x. Now apply Lemmas 1, 3, and 4. Q.E.D.

Proof of Theorem 1:

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

$$\tau_n=\int_0^1 B_n^2(x)dx+o_p(1).$$

The result follows by writing $\int_0^1[B_n^2(x)-\boldsymbol{E}B_n(x)^2]dx$ as a degenerate U statistic of order two. See, for example Serfling (1980, pp. 193-194). 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

$\|\hat{\Omega}-\tilde{\Omega}\| = O[(\log n)/(nh^2)^{1/2}]$. Part (i) of the theorem follows by combining these two results. Part (ii) is an immediate consequence of part (i). Q.E.D.

Proof of Theorem 3: Let $\tilde{z}_\alpha$ denote the $1-\alpha$ quantile of the distribution of $\sum_{j=1}^{\infty}\tilde{\omega}_j \chi_{1j}^2$. Because of Theorem 2, it suffices to show that if 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

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

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

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

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

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

$$S_n(x) = B_n(x) + S_{n2}(x) + S_{n5}(x) - \boldsymbol{E}(W\Delta)'\tilde{\gamma}'(TG_\theta)(x) + o_p(1)$$

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

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

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

almost surely uniformly over x. In addition

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

$$= (T\Delta)(x) + o(1)$$

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

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

where $\tilde{b}_j = b_j + \mu_j$ and b_j is defined as in the proof of Theorem 1. The b_j's are asymptotically distributed as independent $N(\mu_j, \omega_j)$ variates. Now proceed as in Serfling's (1980, pp. 195-199) derivation of the asymptotic distribution of a 2nd-order degenerate U statistic. Q.E.D.

Proof of Theorem 5: Let $z_{g\alpha}$ denote the critical value under the model $Y = g(X) + U$, $g \in \mathcal{F}_{nc}$. Let $\hat{z}_{\varepsilon\alpha g}$ denote the corresponding estimated approximate critical value. Observe that $z_{g\alpha}$ is bounded and $\hat{z}_{\varepsilon\alpha g}$ is bounded in probability uniformly over $g \in \mathcal{F}_{nc}$.

We prove (2.12). The proof of (2.13) is similar. Define $D_n(x) = S_{n3}(x) + S_{n6}(x) + \boldsymbol{E}[S_{n2}(x) + S_{n5}(x)]$ and $\tilde{S}_n(x) = S_n(x) - D_n(x)$. Then $\tau_n = \left\| \tilde{S}_n + D_n \right\|^2$. Use the inequality

(A1) $\quad a^2 \geq 0.5b^2 - (b-a)^2$

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

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

For any finite $M > 0$,

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

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

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

$\left\|\tilde{S}_n\right\|$ 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 $M > M_\varepsilon$

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

Equivalently,

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

and

(A2) $\boldsymbol{P}(\tau_n > z_{g\alpha}) \geq \boldsymbol{P}\left(.5\|D_n\|^2 > z_{g\alpha} + M\right) - \varepsilon$.

Now

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

Therefore,

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

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

$$\boldsymbol{E}[S_{n2}(x) + S_{n5}(x)] = n^{1/2}(Tq_g)(x) + O\left[n^{1/2} h^2 \|q_g\|\right]$$

and

$$\boldsymbol{E}[S_{n2}(x) + S_{n5}(x)] \geq 0.5 n^{1/2}(Tq_g)(x)$$

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

Now

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

Therefore, it follows from the definition $\mathcal{F}_{nc}$ and uniform convergence of $\hat{f}_{XW}^{(-i)}$ to f_{XW} that $\|S_{n3} + S_{n6}\| = O_p(1)$ uniformly over $g \in \mathcal{F}_{nc}$. A further application of (A1) with $a = D_n(x)$ and $b = \boldsymbol{E}[S_{n2}(x) + S_{n5}(x)]$ gives

(A3) $\|D_n\|^2 \geq .125 n \|Tq_g\|^2 + O_p(1)$

uniformly over $g \in \mathcal{F}_{nc}$ as $n \to \infty$. Inequality (2.12) follows by substituting (A3) into (A2) and choosing C to be sufficiently large. Q.E.D.

REFERENCES

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. (1990). A Consistent Conditional Moment Test of Functional Form, *Econometrica*, 58, 1443-1458.

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.

Donald, S.G., G.W. Imbens, and W.K. Newey (2003). Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions, *Journal of Econometrics*, 117, 55-93.

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.

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.

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.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Tripathi, G. and Y. Kitamura (2003). Testing Conditional Moment Restrictions, *Annals of Statistics*, 31, 2059-2095.

Table 1: Results of Monte Carlo Experiments

Null Model	Alt. Model	ρ	η	Empirical Probability that H_0 Is Rejected Using τ_n	t test	Bierens' Test
				H_0 is true		
(4.1)		0.8	0.1	0.051	0.052	0.053
		0.8	0.5	0.030	0.034	0.029
		0.7	0.1	0.049	0.052	0.053
(4.2)		0.8	0.1	0.053	0.040	0.054
		0.8	0.5	0.046	0.077	0.043
		0.7	0.1	0.056	0.036	0.036
				H_0 is false		
(4.1)	(4.2)	0.8	0.1	0.658	0.714	0.470
		0.8	0.5	0.721	0.827	0.466
		0.7	0.1	0.421	0.444	0.280
(4.1)	(4.3)	0.8	0.1	0.684	0.671	0.479
		0.8	0.5	0.663	0.580	0.464
		0.7	0.1	0.424	0.412	0.274
(4.2)	(4.3)	0.8	0.1	0.510	0.566	0.038
		0.8	0.5	0.972	0.987	0.030
		0.7	0.1	0.527	0.590	0.059