A Nonparametric Test of Exogeneity

Richard Blundell
Joel Horowitz

A NONPARAMETRIC TEST OF EXOGENEITY

by

Richard Blundell
Department of Economics
University College London
London WC1E 6BT
UNITED KINGDOM

and

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

November 2004

ABSTRACT

This paper is concerned with inference about a function g that is identified by a conditional moment restriction involving instrumental variables. The function is nonparametric. It satisfies mild regularity conditions but is otherwise unknown. The paper presents test of the hypothesis that g is the mean of a random variable Y conditional on a covariate X. The need to test this hypothesis arises frequently in economics. The test does not require nonparametric instrumental-variables (IV) estimation of g and is not subject to the ill-posed inverse problem that nonparametric IV estimation entails. The test is consistent whenever g differs from the conditional mean function of Y on a set of non-zero probability. Moreover, the power of the test is arbitrarily close to 1 uniformly over a set of functions g whose distance from the conditional mean function is $O(n^{-1/2})$, where n is the sample size.

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

A NONPARAMETRIC TEST OF EXOGENEITY

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

$$(1.1) \quad \boldsymbol{E}[Y - g(X) \,|\, W] = 0 .$$

In (1.1), Y is the dependent variable, X is an explanatory variable, and W is an instrument for X. The function g is nonparametric; it is assumed to satisfy mild regularity conditions but is otherwise unknown. Define the conditional mean function $G(x) = \boldsymbol{E}(Y \,|\, X = x)$. We say that X is exogenous if $g(x) = G(x)$ except, possibly, if x is contained in a set of zero probability. Otherwise, we say that X is endogenous. This paper presents a test of the null hypothesis, H_0, that X is exogenous against the alternative hypothesis, H_1, that X is endogenous. Under mild conditions, the test is consistent whenever $g(x) \neq G(x)$ on a set of non-zero probability. Moreover, in large samples the power of the test is arbitrarily close to 1 uniformly over a set of functions g whose "distance" from G is $O(n^{-1/2})$, where n is the sample size.

The problem of testing exogeneity arises frequently in economics. For example, suppose that Y denotes the hourly wage of a randomly sampled individual and that X includes the individual's level of education among other variables. The random variable $U \equiv Y - g(X)$ normally includes personal characteristics such as "ability" that are not observed by the analyst. If high-ability individuals tend to choose high levels of education, then education is correlated with ability, thereby causing U to be correlated with at least some components of X. When this happens, $g(x) \neq G(x)$, and the precision of any nonparametric estimator of g is typically much lower than that of an estimator of G (Hall and Horowitz 2003). Thus, there is a large loss of estimation efficiency from unnecessarily treating X as endogenous. On the other hand, erroneously assuming that X is exogenous produces a specification error that may be highly misleading. Therefore, it is important have ways to test exogeneity of X. If g is known up to a finite-dimensional parameter, then exogeneity can be tested by using methods developed by Hausman (1978), Bierens (1990), and Bierens and Ploberger (1997). However, these tests can give misleading results if g is misspecified. This paper presents the first test of exogeneity for a nonparametric g.

One possible way to make such a test is to compare nonparametric estimates of g and G. Nonparametric estimators of g have been developed by 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, (1.1) is a Fredholm equation of the first kind, which leads to an ill-posed inverse problem (O'Sullivan 1986, Kress 1999). Consequently, the rate of convergence in probability of a nonparametric estimator of g is typically very slow. Depending on the details of the probability distribution of (Y, X, W), the rate may be slower than $O_p(n^{-\varepsilon})$ for any $\varepsilon > 0$ (Hall and Horowitz 2003). Therefore, a test based on direct comparison of nonparametric estimates of g and G is likely to have low power. Moreover, obtaining the asymptotic distribution of a nonparametric estimator of g is very difficult, and no existing estimator has a known asymptotic distribution. Therefore, it is desirable to avoid nonparametric estimation of g in forming a test of H_0.

The test developed here does not require nonparametric estimation of g and 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 $O_p(n^{-1/2})$. In contrast, the test described in this paper can detect a large class of functions g whose distance from the conditional mean function G in a suitable metric 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, for example, Guerre and Lavergne (2002) and Horowitz and Spokoiny (2001, 2002).

Section 2 of the paper presents the test. Section 3 presents the results of a Monte Carlo investigation of the finite-sample performance of the test. Section 4 presents an economics application that consists of testing the hypothesis that the income variable in an Engel curve is exogenous. 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 a compact set that we take to be $[0,1]^p$ ($p \geq 1$), and Z is a random variable

2

whose support is contained in a compact set that we take to be $[0,1]^r$ ($r \geq 0$). If $r = 0$, then Z is not included in (2.1). We say that 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

(2.2) $\quad \boldsymbol{E}(U \mid X = x, Z = z) = 0$

except, possibly, if (x,z) belongs to a set of probability 0. The alternative hypothesis, H_1, is that (2.2) does not hold on some set $B \subset [0,1]^{p+r}$ that has non-zero probability. 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). Define $G(x,z) = \boldsymbol{E}(Y \mid X = x, Z = z)$. For each $z \in [0,1]^r$, define the operator T_z on $L_2[0,1]^p$ by

$$T_z \psi(x,z) = \int t_z(\xi, x) \psi(\xi, z) d\xi,$$

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) $\quad \tilde{S}(x,z) \equiv T_z(g - G)(x,z) = 0$

for almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that (2.3) does not hold on a set $B \subset [0,1]^{p+r}$ with non-zero Lebesgue measure. 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 $r > 0$. The rate $n^{-1/2}$ 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 $L_2[0,1]^r$. That is, L is defined by

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

and is nonsingular. Define the operator T on $L_2[0,1]^{p+r}$ by $(T\psi)(x,z) = (LT_z)\psi(x,z)$. Then H_0 is equivalent to

(2.4) $\quad S(x,z) \equiv T(g - G)(x,z) = 0$

for almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that (2.4) does not hold on a set $B \subset [0,1]^{p+r}$ with non-zero Lebesgue measure. The test statistic is based on a sample analog of $\int S(x,z)^2 \, dxdz$.

The motivation for basing a test of H_0 on $S(x,z)$ can be understood by observing that $g(x,z) = T_z^{-1} Q(x,z)$, where $Q(x,z) = f_Z(z) E_{W|Z}[E(Y|Z=z,W) f_{XZW}(x,z,W)|Z=z)]$ and f_Z is the probability density function of Z (Hall and Horowitz 2003). T_z^{-1} is a discontinuous operator, and this discontinuity is the source of the ill-posed inverse problem in estimating g. Basing the test of H_0 on $S(x,z)$ avoids this problem because $S(x,z) = L(Q - T_z G)(x,z)$, which does not involve T_z^{-1}.

To form a sample analog of $S(x,z)$, observe that $S(x,z) = E\{[Y - G(X,Z)] \times f_{XW}(x,z,W)\ell(Z,z)\}$. Therefore, the analog can be formed by replacing G and f_{XW} with estimates and E with the sample average in $E\{[Y - G(X,Z)]f_{XW}(x,z,W)\ell(Z,z)\}$. To do this, let $\hat{f}_{XZW}^{(-i)}$ and $\hat{G}^{(-i)}$, respectively, denote leave-observation-i-out "boundary kernel" estimators of f_{XZW} and G. To describe these estimators, let $K_h(\cdot,\cdot)$ denote a boundary kernel function with the property that for all $\xi \in [0,1]$ and some integer $s \geq 2$

(2.5) $\quad h^{-(j+1)} \int_\xi^{\xi+1} u^j K_h(u,\xi) du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq s-1. \end{cases}$

Here, $h > 0$ denotes a bandwidth, and the kernel is defined in generalized form to overcome edge effects. In particular, if h is small and ξ is not close to 0 or 1, then we can set $K_h(u,\xi) = K(u/h)$, where K is an "ordinary" order s kernel. If ξ is close to 1, then we can set $K_h(u,\xi) = \bar{K}(u/h)$, where $\bar{K}$ is a bounded, compactly supported function satisfying

$\int_0^\infty u^j \bar{K}(u) du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq s-1. \end{cases}$

If ξ is close to 0, we can set $K_h(u,\xi) = \bar{K}(-u/h)$. There are, of course, other ways of overcoming the edge-effect problem, but the boundary kernel approach used here works satisfactorily and is simple analytically.

Now define

$K_{p,h}(x,\xi) = \prod_{k=1}^p K_h\left(x^{(k)}, \xi^{(k)}\right),$

where $x^{(k)}$ denotes the k'th component of the vector x. Define $K_{r,h}$ similarly. Then

$$\hat{f}_{XZW}^{(-i)}(x,z,w) =$$

$$\frac{1}{nh_1^{2p+r}}\sum_{\substack{j=1\\j\neq i}}^{n} K_{p,h_1}\left(x-X_j,x\right)K_{p,h_1}\left(w-W_j,w\right)K_{r,h_1}\left(z-Z_j,z\right)$$

and

$$\hat{G}^{(-i)}(x,z) = \frac{1}{nh_2^{p+r}\,\hat{f}_{XZ}^{(-i)}(x,z)}\sum_{\substack{j=1\\j\neq i}}^{n} Y_i K_{p,h_2}\left(x-X_j,x\right)K_{r,h_2}\left(z-Z_j,z\right),$$

where h_1 and h_2 are bandwidths, and

$$\hat{f}_{XZ}^{(-i)}(x,z) = \frac{1}{nh_2^{p+r}}\sum_{\substack{j=1\\j\neq i}}^{n} K_{p,h_2}\left(x-X_j,x\right)K_{r,h_2}\left(z-Z_j,z\right).$$

The sample analog of $S(x,z)$ is

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

The test statistic is

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

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. Let $\left\|(x_1,z_1,w_1)-(x_2,z_2,w_2)\right\|$ denote the Euclidean distance between the points (x_1,z_1,w_1) and (x_2,z_2,w_2) in $[0,1]^{2p+r}$. Let $D_j f_{XZW}$ denote any j'th partial or mixed partial derivative of f_{XZW}. Set $D_0 f_{XZW}(x,z,w) = f_{XZW}(x,z,w)$. Let $s \geq 2$ be an integer. Define $V = Y - G(X,Z)$, and let f_{XZ} denote the density of (X,Z). The assumptions are as follows.

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. (ii) There is a constant $C_X > 0$ such that $f_{XZ}(x,z) \geq C_X$ for all $(x,z) \in \text{supp}(X,Z)$. (iv) 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 $j = 0,1,...,s$, where derivatives

at the boundary of $\text{supp}(X, Z, W)$ are defined as one-sided. (iv) $|D_s f_{XZW}(x_1, z_1, w_1) - D_s f_{XZW}(x_2, z_2, w_2)| \le C_f \|(x_1, z_1, w_1) - (x_2, z_2, w_2)\|$ for any s'th derivative and any $(x_1, z_1, w_1), (x_2, z_2, w_2) \in [0,1]^{2p+r}$. (v) T_z is nonsingular for almost every $z \in [0,1]^r$.

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

3. The conditional mean function G satisfies $|D_j G(x,z)| \le C_f$ for all $(x,z) \in [0,1]^{p+r}$ and $j = 0, 1, ..., s$. (ii) $|D_s G(x_1, z_1) - D_s G(x_2, z_2)| \le C_f \|(x_1, z_1) - (x_2, z_2)\|$ for any s'th derivative and any $(x_1, z_1, x_2, z_2) \in [0,1]^{2(p+r)}$. (iii) $\boldsymbol{E}(V^2 \mid X = x, Z = z) \le C_{UV}$ for each $(x,z) \in [0,1]^{p+r}$.

4. (i) K_h satisfies (2.5) and $|K_h(u_2, \xi) - K_h(u_1, \xi)| \le C_K |u_2 - u_1|/h$ for all u_2, u_1, all $\xi \in [0,1]$, and some constant $C_K < \infty$. For each $\xi \in [0,1]$, $K_h(h, \xi)$ is supported on $[(\xi - 1)/h, \xi/h] \cap \mathcal{K}$, where $\mathcal{K}$ is a compact interval not depending on ξ. Moreover,

$$\sup_{h > 0, \xi \in [0,1], u \in \mathcal{K}} |K_h(hu, \xi)| < \infty.$$

(ii) The bandwidth h_1 satisfies $h_1 = c_{h1} n^{-1/(2s+2p+r)}$, where $c_{h1} < \infty$ is a constant. (iii) The bandwidth, h_2, satisfies $h_2 = c_{h2} n^{-\alpha}$, where $c_{h2} < \infty$ is a constant and $1/(2s) < \alpha < 1/(p+r)$.

Assumption 1(ii) is used to avoid imprecise estimation of G in regions where f_{XZ} is close to 0. The assumption can be relaxed by replacing the fixed distribution of (X, Z, W) by a sequence of distributions with densities $\{f_{nXZW}\}$ and $\{f_{nXZ}\}$ ($n = 1, 2, ...$) that satisfy $f_{nXZ}(x, z) \ge C_n$ for all $(x, z) \in [0,1]^{p+q}$ and a sequence $\{C_n\}$ of strictly positive constants that converges to 0 sufficiently slowly. This complicates the proofs but does not change the results reported here. Assumption 4(iii) implies that the estimator of G is undersmoothed. Undersmoothing prevents the asymptotic bias of $\hat{G}^{(-i)}$ from dominating the asymptotic distribution of τ_n. Assumption 4 requires the use of a higher-order kernel if $p + r \ge 4$. The remaining assumptions are standard in nonparametric estimation.

2.3 *Asymptotic Properties of the Test Statistic*

To obtain the asymptotic distribution of τ_n under H_0, define $V_i = Y_i - G(X_i, Z_i)$

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

and

$$R(x_1,z_1;x_2,z_2) = E[B_n(z_1)B_n(z_2)].$$

Under H_0, $U_i = V_i$. The distinction between U_i and V_i in the definition of B_n will be used later to investigate the distribution of τ_n when H_0 is false. Define the operator Ω on $L_2[0,1]^{p+r}$ by

$$(\Omega\psi)(x,z) = \int_0^1 R(x,z;\xi,\zeta)\psi(\xi,\zeta)d\xi d\zeta .$$

Let $\{\omega_j : j=1,2,...\}$ denote the eigenvalues of Ω sorted so that $\omega_1 \geq \omega_2 \geq ... \geq 0$. 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.

<u>Theorem 1</u>: Let H_0 be true. Then under assumptions 1-4,

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

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. 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 $n \to \infty$. The approximate $1-\alpha$ critical value of the τ_n test is a consistent estimator of the $1-\alpha$ quantile of the approximate distribution.

We now describe the approximation to the asymptotic distribution of τ_n. Under H_0, τ_n is asymptotically distributed as

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

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

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

uniformly over t. Define

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

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. 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 ω_j under H_0. 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 can be estimated with arbitrary accuracy by simulation.

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, this would likely require stronger assumptions than are made here while providing 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 $\tilde{\tau}_{\varepsilon}$ is a complicated function of the spacings and multiplicities of the $\tilde{\omega}_j$'s (Hall and Horowitz 2004). The spacings and multiplicities are unknown in applications and appear difficult to estimate reliably.

The remainder of this section explains how to obtain the estimated eigenvalues $\{\hat{\omega}_j\}$. Because $V = U$ under H_0, a consistent estimator of $R(x_1, z_1; x_2, z_2)$ can be obtained by replacing unknown quantities with estimators on the right-hand side of

$$R(x_1, z_1; x_2, z_2) =$$

$$\boldsymbol{E}\left\{\left[f_{XZW}(x_1, Z, W) - \frac{t_Z(X, x_1)}{f_{XZ}(X, Z)}\right]\left[f_{XZW}(x_2, Z, W) - \frac{t_Z(X, x_2)}{f_{XZ}(X, Z)}\right]\ell(Z, z_1)\ell(Z, z_2)V^2\right\}$$

To do this, let $\hat{f}_{XZW}$ be a kernel estimator of f_{XZW} with bandwidth h. Define

$$\hat{t}_z(x_1, x_2) = \int_0^1 \hat{f}_{XZW}(x_1, z, w)\hat{f}_{XZW}(x_2, z, w)dw .$$

Estimate the V_i 's by

$$\hat{V}_i = Y_i - \hat{G}^{(-i)}(X_i, Z_i).$$

$R(x_1, z_1; x_2, z_2)$ is estimated consistently by

$$\hat{R}(x_1, z_1, x_2, z_2) =$$

$$n^{-1}\sum_{i=1}^{n}\left[\hat{f}_{XZW}(x_1, Z_i, W_i) - \frac{\hat{t}_{Z_i}(X_i, x_1)}{\hat{f}_{XZ}(X_i, Z_i)}\right]\left[\hat{f}_{XZW}(x_2, Z_i, W_i) - \frac{\hat{t}_{Z_i}(X_i, x_2)}{\hat{f}_{XZ}(X_i, Z_i)}\right]\ell(Z_i, z_1)\ell(Z_i, z_2)\hat{V}_i^2.$$

Define the operator $\hat{\Omega}$ on $L_2[0,1]^{p+r}$ by

$$(\hat{\Omega}\psi)(x, z) = \int_0^1 \hat{R}(x, z; \xi, \zeta)\psi(\xi, \zeta)d\xi d\zeta.$$

Denote the eigenvalues of $\hat{\Omega}$ by $\{\hat{\omega}_j : j = 1, 2, ...\}$ and order them so that $\hat{\omega}_{M1} \geq \hat{\omega}_{M2} \geq ... \geq 0$. The relation between the $\hat{\omega}_j$'s and ω_j 's is given by the following theorem.

<u>Theorem 2</u>: Let assumptions 1-4 hold. Then $\hat{\omega}_j - \omega_j = o_p[(\log n)/(nh^{2p+r})^{1/2}]$ as $n \to \infty$, for each $j = 1, 2, ...$

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 $[\hat{f}_{XZW}(x, Z_i, W_i) - \hat{t}_{Z_i}(X_i, x)/\hat{f}_{XW}(X_i, Z_i)]\ell(Z_i, z)$, and let Υ denote the $n \times n$ diagonal matrix whose (i, i) element is $\hat{V}_i^2$. Then

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

The computation of the eigenvalues can now be reduced to finding the eigenvalues of a finite-dimensional matrix. To this end, let $\{\phi_j : j = 1, 2, ...\}$ be a complete, 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

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

and

$$\hat{t}_Z(X, x)\ell(Z, z) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\hat{a}_{jk}\phi_j(x, z)\phi_k(X, Z),$$

where

9

$$\hat{a}_{jk} = \int_0^1 dx_1 \int_0^1 dx_2 \int_0^1 dz_1 \int_0^1 dz_2 \hat{t}_{z_1}(x_1,x_2)\ell(z_1,z_2)\phi_j(x_2,z_2)\phi_k(x_1,z_1).$$

Approximate $\hat{f}_{XZW}(x,z,W)\ell(Z,z)$ and $\hat{t}_z(X,x)\ell(Z,z)$ by the finite sums

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

and

$$\Pi_t(x,z,X,Z)) = \sum_{j=1}^{L}\sum_{k=1}^{L}\hat{a}_{jk}\phi_j(x,z)\phi_k(X,Z).$$

for some integer $L < \infty$. Since $\hat{f}_{XZW}\ell$ and $t_Z\ell$ are known functions, L can be chosen to approximate $\hat{f}_{XZW}\ell$ and $\hat{t}_Z\ell$ with any desired accuracy. Let Φ be the $n \times L$ matrix whose (i,j) component is

$$\Phi_{ij} = n^{-1/2}\sum_{k=1}^{L}[\hat{d}_{jk}\phi_k(Z_i,W_i) - \hat{a}_{jk}\phi_k(X_i,Z_i)/\hat{f}_{XZ}(X_i,Z_i)].$$

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

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

In this section, it is assumed that H_0 is false. That is, $\boldsymbol{P}[X,Z : g(X,Z) = G(X,Z)] < 1$. Define $q(x,z) = g(x,z) - G(x,z)$. Let $\tilde{z}_\alpha$ denote the $1-\alpha$ quantile of the distribution of τ_n under sampling from the null-hypothesis model $Y = G(X,Z) + V$, $\boldsymbol{E}(V \mid X,Z) = 0$. The following theorem establishes consistency of the τ_n test against a fixed alternative hypothesis.

Theorem 3: Suppose that

$$\int_0^1 [(Tq)(x,z)]^2\,dxdz > 0.$$

Let assumptions 1-4 hold. Then for any α such that $0 < \alpha < 1$,

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

Because T is nonsingular, the τ_n test is consistent whenever $g(x,z)$ differs from $G(x,z)$ on a set of (x,z) values whose probability 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

(2.6) $Y = g(X,Z) + U$; $\boldsymbol{E}(U \mid Z,W) = 0$; $\boldsymbol{E}(U \mid X,Z) = n^{-1/2}\Delta(X,Z)$,

where Δ is a bounded function on $[0,1]^{p+r}$. Under (2.6), the distributions of U and V depend

on n, $n^{1/2}(U-V) = \Delta(X,Z)$, and $G(X,Z) = g(X,Z) + n^{-1/2}\Delta(X,Z)$. To provide a complete

characterization of the sequence of alternative hypotheses, it is necessary to specify the

dependence of the distributions of U and V on n. Here, it is assumed that

(2.7) $V = v + n^{-1/2}\varepsilon$,

where ε and v are random variables whose distributions do not depend on n,

$\boldsymbol{E}(v \mid X,Z) = \boldsymbol{E}(v \mid Z,W) = 0$, $Var(v) < \infty$, $\boldsymbol{E}(\varepsilon \mid X,Z) = 0$, $\boldsymbol{E}(\varepsilon \mid Z,W) = -\boldsymbol{E}[\Delta(X,Z) \mid Z,W]$,

and $Var(\varepsilon) < \infty$. It follows from (2.6)-(2.7) that

(2.8) $U = v + n^{-1/2}\Delta(X,Z) + n^{-1/2}\varepsilon$,

The following additional notation is used. Define

$$\tilde{B}_n(x,z) = n^{-1/2}\sum_{i=1}^{n} v_i[f_{XZW}(x,Z_i,W_i) - t_{Z_i}(X_i,x)/f_{XZ}(X_i,Z_i)]\ell(Z_i,z)$$

and $\tilde{R}(x_1,z_1;x_2,z_2) = \boldsymbol{E}[\tilde{B}_n(x_1,z_1)\tilde{B}_n(x_2,z_2)]$. Define the operator $\tilde{\Omega}$ on $L_2[0,1]^{p+r}$ by

$$(\tilde{\Omega}\psi)(x,z) = \int_0^1 \tilde{R}(x,z;\xi,\zeta)\psi(\xi,\zeta)d\xi d\zeta .$$

Let $\{(\tilde{\omega}_j, \psi_j): j = 1,2,...\}$ denote the eigenvectors and orthonormal eigenvectors of $\tilde{\Omega}$. Define

$\mu(x,z) = (T\Delta)(x,z)$ and

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

Let $\{\chi_{1j}^2(\mu_j^2/\tilde{\omega}_j): j = 1,2,...\}$ denote independent random variables that are distributed as non-

central chi-square with one degree of freedom and non-central parameters $\{\mu_j^2/\tilde{\omega}_j\}$. The

following theorem states the result.

<u>Theorem 4</u>: Let assumptions 1-4 hold. Under the sequence of local alternatives (2.6)-
(2.8),

$$\tau_n \to^d \sum_{j-1}^{\infty} \tilde{\omega}_j \chi_{1j}^2(\mu_j^2/\tilde{\omega}_j) .$$

It follows from Theorems 2 and 4 that under (2.6)-(2.8),

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

for any $\varepsilon > 0$, where $\hat{z}_{\varepsilon\alpha}$ denotes the estimated approximate α-level critical value.

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 set of functions g whose distance from G is $O(n^{-1/2})$. The following additional notation is used. Define $q(x,z) = g(x,z) - G(x,z)$. Let f_{XZW} be fixed. For each $n = 1, 2, \ldots$ and finite $C > 0$, define $\mathcal{F}_{nc}$ as a set of distributions of (Y, X, Z, W) such that: (i) f_{XZW} satisfies assumption 1; (ii) $\boldsymbol{E}[Y - g(X,Z) | Z, W] = 0$ for some function g that satisfies assumption 2 with $U = Y - g(X,Z)$; (iii) $\boldsymbol{E}[Y - G(X,Z) | X, Z] = 0$ for some function G that satisfies assumption 3 with $V = Y - G(X,Z)$; (iv) $\|Tq\| \geq n^{-1/2} C$, where $\|\cdot\|$ denotes the L_2 norm; and (v) $h_1^s (\log n) \|q\| / \|Tq\| = o(1)$ as $n \to \infty$. $\mathcal{F}_{nc}$ is a set of distributions of (Y, X, Z, W) for which the distance of g from G shrinks to zero at the rate $n^{-1/2}$ in the sense that $\mathcal{F}_{nc}$ includes distributions for which $\|q\| = O(n^{-1/2})$. Condition (v) rules out distributions for which q depends on (x,z) only through sequences of eigenvectors of T whose eigenvalues converge to 0 too rapidly. For example, let $p = 1$, $r = 0$, so Z is not in the model. Let $\{\lambda_j, \phi_j : j = 1, 2, \ldots\}$ denote the eigenvalues and eigenvectors of T ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots > 0$. Suppose that $G(x) = \phi_1(x)$, $g(x) = \phi_1(x) + \phi_n(x)$, and the instrument is $\tilde{W} = \phi_1(W)$. Then $h_1^2 \|q\| / \|Tq\| = h_1^2 / \lambda_n$. Because $h_1 \propto n^{-1/6}$, condition (v) is violated if $\lambda_n = o(n^{-1/3})$. The practical significance of condition (v) is that the τ_n test has low power when g differs from G only through eigenvectors of T with very small eigenvalues.

The following theorem states the result of this section.

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

$$\liminf_{n \to \infty \ \mathcal{F}_{nc}} \boldsymbol{P}(\tau_n > z_\alpha) \geq 1 - \delta$$

and

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

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 $H(\cdot, \cdot)$ be a bounded, real-valued function on $[0,1]^2$ with the property that

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

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

$$\tau_{nH} = \int_0^1 S_{nH}^2(z)dz ,$$

where

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

If $H(z,w) = \tilde{H}(zw)$ 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 f_{XW} and local alternative models (2.6)-(2.8) such that an α-level τ_{nH} test based on a fixed H that does not depend on the sampled population 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 G is known and does not have to be estimated. Define

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

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

$\bar{R}(z_1, z_2) = E[\bar{B}_n(z_1)\bar{B}_n(z_2)]$, and $\bar{R}_H(z_1, z_2) = E[\bar{B}_{nH}(z_1)\bar{B}_{nH}(z_2)]$. Also, define the operators $\bar{\Omega}$ and $\bar{\Omega}_H$ on $L_2[0,1]$ by

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

and

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

Let $\{\bar{\omega}_j, \bar{\psi}_j : j = 1,2,...\}$ and $\{\bar{\omega}_{jH}, \bar{\psi}_{jH} : j = 1,2,...\}$ 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 (2.6), define $\bar{\mu}(z) = (T\Delta)(z)$,

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

$$\bar{\mu}_j = \int_0^1 \bar{\mu}(z)\bar{\psi}_j(z)dz ,$$

and

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

Then arguments like those used to prove Theorem 4 show that under the sequence of local alternatives (2.6)-(2.8) with a known function G,

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

and

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

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 $\|\bar{\mu}\|^2 / \sum_{j=1}^{\infty}\bar{\omega}_j$ is bounded away from 0 and $\|\bar{\mu}_H\|^2 / \sum_{j=1}^{\infty}\bar{\omega}_{jH}$ 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 $j \geq 1$. Let $\ell > 1$ be a finite integer. Define

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

Let

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

Let $E(U^2 \mid W = w) = 1$ for all $w \in [0,1]$. Then $\bar{R}(z_1, z_2) = t(z_1, z_2)$, $\bar{\omega}_j = \lambda_j$, and $\sum_{j=1}^{\infty} \bar{\omega}_j$ is non-zero and finite. Set $\Delta(x) = D\phi_\ell(x)$ for some finite $D > 0$. Then $\|\bar{\mu}\|^2 = D^2 \lambda_\ell^2 = D^2$. Since H is fixed, it suffices to show that ℓ 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

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

where $\{h_{jk} : j, k = 1, 2, ...\}$ are constants. Moreover, $\|\bar{\mu}_H\|^2 = D^2 \sum_{j=1}^{\infty} h_{j\ell}^2$. Since H is bounded, ℓ can be chosen so that $\sum_{j=1}^{\infty} h_{j\ell}^2 < \varepsilon/D^2$ for any $\varepsilon > 0$. With this ℓ, $\|\bar{\mu}_H\|^2 < \varepsilon$, 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 $U[0,1]$, $E(U^2 \mid W = w) = 1$ 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,

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

It follows from the Cauchy-Schwartz inequality that

$$\|\bar{\mu}_H\|^2 \le \left[\int_0^1 \int_0^1 H(z,w)^2 \, dz\,dw \right] \int_0^1 \left[\int_0 f_{XW}(x,w)\Delta(x)dx \right]^2 dw$$

$$= \int_0^1 \left[\int_0 f_{XW}(x,w)\Delta(x)dx \right]^2 dw$$

$$\le \|\Delta\|^2 \|T\Delta\|^2$$

$$\le C_\Delta \|\bar{\mu}\|^2 .$$

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. In the experiments, $p = 1$ and $r = 0$, so Z does not enter the model. Realizations of (X, W) were generated by $X = \Phi(\xi)$ and $W = \Phi(\zeta)$, 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 $\rho = 0.7$. Realizations of Y were generated from

$$(3.1) \qquad Y = \theta_0 + \theta_1 X + \sigma_U U,$$

where $\theta_0 = 0$, $\theta_1 = 0.5$, $U = \eta\varepsilon + (1-\eta^2)^{1/2}v$, $v \sim N(0,1)$, $\sigma_U = 0.2$, and η is a constant parameter whose value varies among experiments. H_0 is true if $\eta = 0$ and false otherwise. To provide a basis for judging whether the power of the τ_n test is high or low, we also report the results of a Hausman (1978) type test of the hypothesis that the ordinary least squares and instrumental variables (IV) estimators of θ_1 in (3.1) are equal. The instruments used for IV estimation of (3.1) are $(1, W)$. In addition, we report the results of simulations with τ_{nH}. The weight function is $H(x, w) = \exp(xw)$ and is taken from Bierens (1990). The bandwidth used to estimate f_{XW} was selected by cross-validation. The bandwidth used to estimate f_X is $n^{1/5 - 7/24}$ times the cross-validation bandwidth. 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, and the estimated eigenvalues $\hat{\omega}_j$ are very close to 0 when $j > 25$. 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 of the τ_n and Hausman-type tests are small when H_0 is true. When H_0 is false, the power of the τ_n test is, not surprisingly, somewhat smaller than the power of the Hausman-type test, which is parametric, but the differences in power are not great. The performance of τ_{nH} is worse than that of τ_n. When H_0 is true, difference between the nominal and empirical rejection probabilities of the τ_{hH} test is relatively large, and the power of the τ_{nH} test is lower than that of the τ_n test.

4. AN EMPIRICAL EXAMPLE

This section presents an empirical example in which τ_n is used to test the hypothesis that the right-hand side variable of an Engel curve is exogenous. The curve is given by (2.1) with $p = 1$ and $r = 0$, where Y denotes the expenditure share of services, X denotes the logarithm of total expenditures, and W denotes annual income from wages and salaries of the head of household. Engel curves are important in the analysis of consumer behavior. Parametric specifications are often linear or quadratic in X (Muellbauer 1976; Banks, Blundell, and Lewbel 1998). There is also increasing use of nonparametric estimators (Deaton 1998). However, X is arguably jointly determined with household budgeting decisions and, therefore, endogenous.

The data consist of 1518 household-level observations from the British Family Expenditure Survey. This is a diary-based household survey that is supplemented by recall information. We use a subsample consisting of married couples with one or two children and an employed head of household. W should be a good instrument for X if income from wages and salaries is not influenced by household budgeting decisions.

The bandwidths for estimating f_{XW} were selected by the method described in the Monte Carlo section. 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 X is exogenous gives $\tau_n = 0.162$ with a 0.05-level critical value of 0.151. Thus, the test rejects the hypothesis that X is exogenous. The hypothesis was also tested by comparing the OLS and IV estimates of θ_1 and θ_2 in the quadratic model

$$Y = \theta_0 + \theta_1 X + \theta_2 X^2 + U.$$

The instruments are $(1, W, W^2)$. The hypothesis that the OLS estimates of θ_1 and θ_2 equal the IV estimates is rejected at the 0.05 level. Thus, the τ_n test and the parametric test both reject the hypothesis that the logarithm of total expenditures is exogenous.

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

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

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

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

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

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

and

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

Then

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

Define $V_i = Y_i - G(X_i)$.

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

$$S_{n3}(z) = n^{-1/2} \sum_{i=1}^{n} V_i t(X_i, z) / f_X(X_i) + r_n(z),$$

where $\int_0^1 r_n^2(z) dz = o_p(1)$.

Proof: Define

$$R_{n1}^{(-i)}(x) = \frac{1}{n h_2 f_X(x)} \sum_{\substack{j=1 \\ j \neq i}}^{n} V_j K_{h_2}(x - X_j, x),$$

$$R_{n2}^{(-i)}(x) = \frac{1}{n h_2 f_X(x)} \sum_{\substack{j=1 \\ j \neq i}}^{n} [G(X_j) - G(x)] K_{h_2}(x - X_j, x),$$

$$S_{n3a}(z) = n^{-1/2} \sum_{i=1}^{n} E_i [R_{n1}^{(-i)}(X_i) f_{XW}(z, W_i)]$$

where E_i denotes the expected value over i-subscripted random variables,

$$S_{n3b}(z) = n^{-1/2} \sum_{i=1}^{n} \{R_{n1}^{(-i)}(X_i) f_{XW}(z, W_i) - E_i[R_{n1}^{(-i)}(X_i) f_{XW}(z, W_i)]\},$$

and

$$S_{n3c}(z) = n^{-1/2} \sum_{i=1}^{n} R_{n2}^{(-i)}(X_i) f_{XW}(z, W_i)$$

Standard calculations for kernel estimators show that

$$\hat{G}^{(-i)}(x) - G(x) = \frac{1}{nh_2 f_X(x)} \sum_{\substack{j=1 \\ j \neq i}} [Y_j - G(x)] K_{h_2}(x - X_j, x) + O\left[\frac{(\log n)^2}{nh_2} + h_2^4\right]$$

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

$$S_{n3}(z) = -[S_{n3a}(z) + S_{n3b}(z) + S_{n3c}(z)] + o_p(1)$$

uniformly over $z \in [0,1]$. Lengthy but straightforward calculations show that

$$E \int_0^1 S_{n3b}^2(z) dz = o(1), \quad E \int_0^1 S_{n3c}^2(z) dz = o(1)$$

as $n \to \infty$. Therefore,

(A.1) $\int_0^1 S_{n3b}^2(z) dz = o_p(1)$

and

(A.2) $\int_0^1 S_{n3c}^2(z) dz = o_p(1)$

by Markov's inequality. Moreover, we can write

$$E_i[R_{n1}^{(-i)}(X_i) f_{XW}(z, W_i)]$$

$$= \frac{1}{nh_2} \sum_{\substack{j=1 \\ j \neq i}}^{n} V_j \int_0^1 [f_{XW}(x, w) f_{XW}(z, w) / f_X(x)] K_{h_2}(x - X_j, x) dx dw$$

$$= \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} V_j [t(X_j, z) / f_X(X_j) + \rho_{n1}(X_j, z)],$$

where $\rho_{n1}(x, z) = O(h_2^2)$ uniformly over $(x, z) \in [0,1]^2$. Therefore,

(A.3) $S_{n3a}(z) = n^{-1/2} \sum_{i=1}^{n} V_i t(X_i, z) / f_X(X_i) + \rho_{n2}(z)$

where $E \int_0^1 \rho_{n2}^2(z) dz = o(1)$ as $n \to \infty$. The lemma follows by combining (A.1)-(A.3). Q.E.D.

<u>Lemma 2</u>: As $n \to \infty$, $\int_0^1 S_{n4}^2(z)dz = o_p(1)$.

Proof: Define

$$D_n = n^{-1}\boldsymbol{E}\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n U_i U_j \int_0^1 [\hat{f}_{XW}^{(-i)}(z,W_i) - f_{XW}(z,W_i)][\hat{f}_{XW}^{(-j)}(z,W_i) - f_{XW}(z,W_j)]dz .$$

Then

$$\boldsymbol{E}\int_0^1 S_{n4}^2(z)dz = D_n + n^{-1}\boldsymbol{E}\sum_{i=1}^n U_i^2 \int_0^1 [\hat{f}_X^{(-i)}(z,W_i) - f_X(z,W_i)]^2 dz$$

(A.4)
$$= D_n + o(1).$$

Now define

$$\hat{f}_{XW}^{(-i,-j)}(z,w) = \frac{1}{nh_1^2}\sum_{\substack{k=1 \\ k \neq i,j}}^n K_{h_1}(z - X_k, z)K_{h_1}(w - W_k, w)$$

and

$$\delta_j(z,w) = \frac{1}{nh_1^2}K_{h_1}(z - X_j, z)K_{h_1}(w - W_j, w).$$

Then $D_n = D_{n1} + 2D_{n2} + D_{n3}$, where

$$D_{n1} =$$

$$n^{-1}\boldsymbol{E}\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n U_i U_j \int_0^1 [\hat{f}_{XW}^{(-i,-j)}(z,W_i) - f_{XW}(z,W_i)][\hat{f}_{XW}^{(-j,-i)}(z,W_i) - f_{XW}(z,W_j)]dz$$

$$D_{n2} = n^{-1}\boldsymbol{E}\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n U_i U_j \int_0^1 [\hat{f}_{XW}^{(-i,-j)}(z,W_i) - f_{XW}(z,W_i)]\delta_j(z,,W_i)dz ,$$

and

$$D_{n3} = n^{-1}\boldsymbol{E}\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n U_i U_j \int_0^1 \delta_i(z,W_j)\delta_j(z,,W_i)dz .$$

But $\boldsymbol{E}(U \mid W) = 0$. Therefore, $D_{n1} = D_{n2} = 0$, and $D_{n3} = O[(nh_1^2)^{-1}]$. The lemma now follows from Markov's inequality. Q.E.D.

<u>Lemma 3</u>: As $n \to \infty$, $S_{n6}(z) = o_p(1)$ uniformly over $z \in [0,1]$.

<u>Proof</u>: This follows from $\hat{f}_{XW}^{(-i)}(x,w) - f_{XW}(x,w) = O[(\log n)/(nh_1^2)^{1/2} + h_1^2]$ almost surely uniformly over $(x,w) \in [0,1]^2$ and $\hat{G}^{(-i)}(x) - G(x) = O[(\log n)/(nh_2)^{1/2} + h_2^2]$ almost surely uniformly over $x \in [0,1]$. Q.E.D.

<u>Proof of Theorem 1</u>: Under H_0, $S_{n2}(z) = S_{n5}(z) = 0$ for all $z \in [0,1]$. Therefore, it follows from Lemmas 1-3 that

$$\tau_n = \int_0^1 B_n^2(z)dz + o_p(1).$$

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

<u>Proof of Theorem 2</u>: $|\hat{\omega}_j - \tilde{\omega}_j| = O(\|\hat{\Omega} - \tilde{\Omega}\|)$ by Theorem 5.1a of Bhatia, Davis, and McIntosh (1983). Moreover, standard calculations for kernel density estimators show that $\|\hat{\Omega} - \tilde{\Omega}\| = O[(\log n)/(nh_1^2)^{1/2}]$. Part (i) of the theorem follows by combining these two results. Part (ii) is an immediate consequence of part (i). Q.E.D.

<u>Proof of Theorem 3</u>: 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

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

To do this, observe that by a uniform law of large numbers of Pakes and Pollard (1989, Lemma 2.8), $n^{-1/2}S_{n2}(z) = (Tq)(z) + o_p(1)$ uniformly over $z \in [0,1]$. Moreover, $n^{-1/2}S_{n5}(z) = o_p(1)$ uniformly over $z \in [0,1]$ because $\hat{f}_{XW}^{(-i)}(z,w) - f_{XW}(z,w) = O[(\log n)/(nh_1^2)^{1/2} + h_1^2]$ a.s. uniformly over $(z,w) \in [0,1]^2$. Combining these results with Lemmas 1-3 yields

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

where $\int_0^1 r_n^2(z)dz = o_p(1)$ as $n \to \infty$. It follows from Theorem 1 that $n^{-1}\int_0^1 B_n^2(z)dz = o_p(1)$.

Therefore, $n^{-1}\tau_n \to^p \int_0^1 [(Tq)(z)]^2 dz$. Q.E.D.

Proof of Theorem 4: The conclusions of lemmas 1-3 hold under (2.6)-(2.8). Therefore,

$$S_n(z) = B_n(z) + S_{n2}(z) + S_{n5}(z) + r_n(z),$$

where $\int_0^1 r_n^2(z)dz = o_p(1)$. Moreover,

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

$$= o(1)$$

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, $S_n(z) = B_n(z) + \mu(z) + r_n(z)$. But

$$B_n(z) = \tilde{B}_n(z) + o_p(1)$$

uniformly over $z \in [0,1]$. Therefore, it suffices to find the asymptotic distribution of

$$\int_0^1 [\tilde{B}_n(z) + \mu(z)]^2 dz = \sum_{j=1}^\infty (\tilde{b}_j + \mu_j)^2,$$

where

$$\tilde{b}_j = \int_0^1 \tilde{B}_n(z)\psi_j(z)dz.$$

The random variables $\tilde{b}_j + \mu_j$ are asymptotically distributed as independent $N(\mu_j, \tilde{\omega}_j)$ variates. Now proceed as in , for example, Serfling's (1980, pp. 195-199) derivation of the asymptotic distribution of a degenerate, order-2 U statistic. Q.E.D.

The following definitions are used in the proof of Theorem 5. For each distribution $\pi \in \mathcal{F}_{nc}$, let $A(\pi)$ be a random variable. Let $\{c_n : n = 1, 2, ...\}$ be a sequence of positive constants. Write $A = O_p(c_n)$ uniformly over $\mathcal{F}_{nc}$ if for each $\varepsilon > 0$ there is a constant M_ε such that

$$\sup_{\pi \in \mathcal{F}_{nc}} P[|A(\pi)|/c_n > M_\varepsilon] < \varepsilon.$$

22

For each $\pi \in \mathcal{F}_{nc}$, let $\{A_n(\pi): n = 1, 2, ...\}$ be a sequence of random variables. Write $A_n = o_p(1)$ uniformly over $\mathcal{F}_{nc}$ if for each $\varepsilon > 0$

$$\lim_{n \to \infty} \sup_{\pi \in \mathcal{F}_{nc}} P[|A_n(\pi)| > \varepsilon] = 0.$$

<u>Proof of Theorem 5</u>: Let z_α denote the critical value of τ_n. Observe that z_α is bounded uniformly over $\mathcal{F}_{nc}$. The arguments used to prove lemmas 1-3 show that $\int_0^1 S_{nj}^2(z)dz = o_p(1)$ for $j = 4, 6$ and $\int_0^1 S_{n3}^2(z)dz = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. In addition, an application of Markov's inequality shows that $\int_0^1 S_{n1}^2(z)dz = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. Define

$$\tilde{S}_n(z) = S_{n1}(z) + S_{n3}(z) + S_{n4}(z) + S_{n6}(z)$$

and

$$D_n(z) = S_{n2}(z) + S_{n5}(z).$$

Let $\|\cdot\|$ denote the $L_2[0,1]$ norm. Use the inequality $a^2 \geq 0.5b^2 - (b-a)^2$ with $a = S_n$ and $b = S_{n2} + S_{n5}$ to obtain

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

For any finite $M > 0$,

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

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

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

$\|\tilde{S}_n\| = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. Therefore, for each $\varepsilon > 0$ there is $M_\varepsilon < \infty$ such that for all $M > M_\varepsilon$

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

for all distributions in $\mathcal{F}_{nc}$. Equivalently,

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

23

and

(A8) $\quad P(\tau_n > z_\alpha) \ge P\left(.5\|D_n\|^2 > z_\alpha + M\right) - \varepsilon$.

Now

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

Therefore,

$$\boldsymbol{E} D_n(z) = n^{-1/2} \boldsymbol{E} \sum_{i=1}^{n} [g(X_i) - G(X_i)][f_{XW}(z, W_i) + h_1^2 R_n(z)] ,$$

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

$$\boldsymbol{E} D_n(z) = n^{1/2} (Tq)(z) + O\left[n^{1/2} h_1^2 \|q\| \right]$$

and

$$\boldsymbol{E} D_n(z) \ge 0.5 n^{1/2} (Tq)(z)$$

for all distributions in $\mathcal{F}_{nc}$ and all sufficiently large n. Moreover,

$$D_n(z) - \boldsymbol{E} D_n(z) = n^{-1/2} \sum_{i=1}^{n} [q(X_i) \boldsymbol{E}^{(-i)} \hat{f}_{XW}^{(-i)}(z, W_i) - \boldsymbol{E} q(X) \hat{f}_{XW}^{(-i)}(z, W_i)]$$

$$+ n^{-1/2} \sum_{i=1}^{n} q(X_i) [f_{XW}^{(-i)}(z, W_i) - \boldsymbol{E}^{(-i)} \hat{f}_{XW}^{(-i)}(z, W_i)]$$

$$\equiv D_{n1}(z) + D_{n2}(z),$$

where $\boldsymbol{E}^{(-i)}$ denotes the expectation with respect to the distribution of $\{X_j, W_j : j = 1,...,n; \ j \ne i\}$. It is clear that $\|D_{n1}\|^2 = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. Moreover, it follows from the properties of kernel estimators that

$$|D_{n2}(z)| \le \frac{r_n \log n}{n h_1} \sum_{n \ i=1}^{n} |q(X_i)|$$

$$= \frac{r_n \log n}{h_1} [\boldsymbol{E} |q(X)| + O_p(n^{-1/2})],$$

uniformly over $\mathcal{F}_{nc}$, where $r_n = O(1)$ almost surely as $n \to \infty$ and depends only on the distribution of (X, W). Therefore,

$$\left\| D_n - \boldsymbol{E} D_n \right\|^2 \le \left(\frac{r_n \log n}{n^{1/2} h_1} \right)^2 n(\boldsymbol{E} \mid q \mid)^2 O_p(1) + O_p(1).$$

A further application of $a^2 \ge 0.5 b^2 - (b-a)^2$ with $a = D_n$ and $b = \boldsymbol{E} D_n$ gives

$$\left\| D_n \right\|^2 \ge n \left\| Tq \right\|^2 \left[.125 - \left(\frac{r_n \log n}{n^{1/2} h_1^3} \right)^2 \frac{h_1^4 (\boldsymbol{E} \mid q \mid)^2}{\left\| Tq \right\|^2} O_p(1) \right] + O_p(1)$$

$$= n \left\| Tq \right\|^2 \left[.125 - \left(\frac{r_n \log n}{n^{1/2} h_1^3} \right)^2 o_p(1) \right] + O_p(1)$$

uniformly over $\mathcal{F}_{nc}$. Therefore, if C is sufficiently large, $0.5 \left\| D_n \right\|^2 > z_\alpha + M$ with probability approaching 1 as $n \to \infty$ uniformly over $\mathcal{F}_{nc}$. Q.E.D.

REFERENCES

Banks, J., R.W. Blundell, and A. Lewbel (1997). Quadratic Engel Curves, Indirect Tax Reform and Welfare Measurement, *Review of Economics and Statistics*, 79, 527-539.

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.

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

Deaton, A. (1998). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

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.

Hall, P. and J.L. Horowitz (2004). Methodology and Convergence Rates for Functional Linear Regression, working paper, Department of Economics, Northwestern University.

Hausman, J.A. (1978). Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

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.

Meullbauer, J. (1976). Community Preferences and the Representative Consumer, *Econometrica*, 44, 525-543.

Kress, R. (1999). *Linear Integral Equations*, 2nd ed., New York: Springer.

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.

O'Sullivan, F. (1986). A Statistical Perspective on Ill-Posed Problems, *Statistical Science*, 1, 502-527.

Pakes, A. and D. Pollard (1988). Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, 57, 1027-1057.

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

Table 1: Results of Monte Carlo Experiments

η	τ_n	Empirical Probability that H_0 Is Rejected Using Hausman test	τ_{nH}
0.0	0.048	0.055	0.025
0.1	0.256	0.304	0.187
0.15	0.539	0.590	0.429
0.20	0.814	0.876	0.724
0.25	0.945	0.971	0.922