cemmap

centre for microdata methods and practice

Testing exogeneity in nonparametric instrumental variables identified by conditional quantile restrictions

Jia-Young Michael Fu
Joel Horowitz
Matthias Parey

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP68/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

TESTING EXOGENEITY IN NONPARAMETRIC INSTRUMENTAL VARIABLES MODELS IDENTIFIED BY CONDITIONAL QUANTILE RESTRICTIONS

by

Jia-Young Michael Fu
Department of Economics
Northwestern University
Evanston, IL 60201

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60201

Matthias Parey
Department of Economics
University of Essex
Colchester CO4 3SQ
United Kingdom

October 2015

Abstract

This paper presents a test for exogeneity of explanatory variables in a nonparametric instrumental variables (IV) model whose structural function is identified through a conditional quantile restriction. Quantile regression models are increasingly important in applied econometrics. As with mean-regression models, an erroneous assumption that the explanatory variables in a quantile regression model are exogenous can lead to highly misleading results. In addition, a test of exogeneity based on an incorrectly specified parametric model can produce misleading results. This paper presents a test of exogeneity that does not assume the structural function belongs to a known finite-dimensional parametric family and does not require nonparametric estimation of this function. The latter property is important because, owing to the ill-posed inverse problem, a test based on a nonparametric estimator of the structural function has low power. The test presented here is consistent whenever the structural function differs from the conditional quantile function on a set of non-zero probability. The test has non-trivial power uniformly over a large class of structural functions that differ from the conditional quantile function by $O(n^{-1/2})$. The results of Monte Carlo experiments illustrate the usefulness of the test.

Key words: Hypothesis test, instrumental variables, quantile estimation, specification testing

JEL Listing: C12, C14

We thank Richard Blundell for helpful comments. Part of this research was carried out while Joel L. Horowitz was a visitor at the Department of Economics, University College London, and the Centre for Microdata Methods and Practice.

TESTING EXOGENEITY IN NONPARAMETRIC INSTRUMENTAL VARIABLES MODELS IDENTIFIED BY CONDITIONAL QUANTILE RESTRICTIONS

1. INTRODUCTION

Econometric models often contain explanatory variables that may be endogenous. For example, in a wage equation, the observed level of education may be correlated with unobserved ability, thereby causing education to be an endogenous explanatory variable. It is well known that estimation methods for models in which all explanatory variables are exogenous do not yield consistent parameter estimates when one or more explanatory variables are endogenous. For example, ordinary least squares does not provide consistent estimates of the parameters of a linear model when one or more explanatory variables are endogenous. Instrumental variables estimation is a standard method for obtaining consistent estimates.

The problem of endogeneity is especially serious in nonparametric estimation. Because of the ill-posed inverse problem, nonparametric instrumental variables estimators are typically much less precise than nonparametric estimators in the exogenous case. Therefore, it is especially useful to have methods for testing the hypothesis of exogeneity in nonparametric settings. This paper presents a test of the hypothesis of exogeneity of the explanatory variable in a nonparametric quantile regression model.

Quantile models are increasingly important in applied econometrics. Koenker (2005) and references therein describe methods for and applications of quantile regression when the explanatory variables are exogenous. Estimators and applications of linear quantile regression models with endogenous explanatory variables are described by Amemiya (1982), Powell (1983), Chen and Portnoy (1996), Januszewski (2002), Chernozhukov and Hansen (2004, 2006), Ma and Koenker (2006), Blundell and Powell (2007), Lee (2007), and Sakata (2007). Nonparametric methods for quantile regression models are discussed by Chesher (2003, 2005, 2007); Chernozhukov and Hansen (2004, 2005, 2006); Chernozhukov, Imbens, and Newey (2007); Horowitz and Lee (2007); and Chen and Pouzo (2009, 2012). Blundell, Horowitz, and Parey (2015) estimate a nonparametric quantile regression model of demand under the hypothesis that price is exogenous and an instrumental variables quantile regression model under the hypothesis that price is endogenous.

The method presented in this paper consists of testing the conditional moment restriction that defines the null hypothesis of exogeneity in a quantile IV model. This approach does not require estimation of the structural function. An alternative approach is to compare a nonparametric quantile estimate of the structural function under exogeneity with an estimate obtained by using nonparametric instrumental variables methods. However, the moment condition that identifies the structural function in the presence of endogeneity is a nonlinear integral equation of the first kind, which leads to an ill-posed inverse problem (O'Sullivan 1986, Kress 1999). A consequence of this is that in the presence of one or more endogenous explanatory variables, the rate of convergence of a nonparametric estimator of the structural function is typically very slow. Therefore, a test based on a direct comparison of nonparametric estimates obtained with and without assuming exogeneity will have low power. Accordingly, it is desirable to have a test of exogeneity that avoids nonparametric instrumental variables estimation of the structural function. This paper presents such a test.

Breunig (2015) and Blundell and Horowitz (2007) have developed tests of exogeneity of the explanatory variables in a nonparametric instrumental variables model that is identified through a conditional mean restriction. The test presented here uses ideas and has properties similar to those of Blundell's and Horowitz's (2007) test. However, the non-smoothness of quantile estimators presents technical issues that are different from and more complicated than those presented by instrumental variables models that are identified by conditional mean restrictions. Therefore, testing exogeneity in a quantile regression model requires a separate treatment from testing exogeneity in the conditional mean models considered by Breunig (2015) and Blundell and Horowitz (2007). We use empirical process methods to deal with the non-smoothness of quantile estimators. Such methods are not needed for testing exogeneity in conditional mean models.

Section 2 of this paper presents the model, null hypothesis to be tested, and test statistic. Section 3 describes the asymptotic properties of the test and explains how to compute the critical value in applications. Section 4 presents the results of a Monte Carlo investigation of the finite-sample performance of the test. Section 5 concludes. The proofs of theorems are in the appendix, which is Section 6.

2. THE MODEL, NULL HYPOTHESIS, AND TEST STATISTIC

This section begins by presenting the model setting that we deal with, the null hypothesis to be tested, and issues that are involved in testing the null hypothesis. Section 2.2 presents the test statistic.

2.1 *The Model and the Null and Alternative Hypotheses*

Let Y be a scalar random variable, X and W be continuously distributed random scalars or vectors, q be a constant satisfying $0 < q < 1$, and g be a structural function that is identified by the relation

(2.1) $$P[Y - g(X) \le 0 \mid W = w] = q$$

for almost every $w \in \operatorname{supp}(W)$. Equivalently, g is identified by

(2.2) $$Y = g(X) + U; \quad P(U \le 0 \mid W = w) = q$$

for almost every $w \in \text{supp}(W)$. In (2.1) and (2.2), Y is the dependent variable, X is the 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 q-quantile function $G(x) = Q_q(Y \mid X = x)$, where Q_q denotes the conditional q-quantile. 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. It follows from (2.1) and (2.2) that H_0 is equivalent to testing the hypothesis $P[g(X) = G(X)] = 1$ or $P[Y - G(X) \le 0 \mid W = w] = q$ for almost every $w \in \text{supp}(W)$. H_1 is equivalent to $P[g(X) = G(X)] < 1$. Under mild conditions, the test presented here rejects H_0 with probability approaching 1 as the sample size increases whenever $g(x) \ne G(x)$ on a set of non-zero probability.

One possible way of testing H_0 is to estimate g and G, compute the difference between the two estimates in some metric, and reject H_0 if the difference is too large. To see why this approach is unattractive, assume that $\text{supp}(X, W) \subset [0,1]^2$. This assumption entails no loss of generality if X and W are scalars. It can always be satisfied by, if necessary, carrying out monotone increasing transformations of X and W. Then (2.1) is equivalent to the nonlinear integral equation

(2.3) $$\int_0^1 F_{YXW}[g(x), x, w]dx - qf_W(w) = 0,$$

where f_W is the probability density function of w,

$$F_{YXW}(y, x, w) = \int_0^y f_{YXW}(u, x, w)du,$$

and f_{YXW} is the probability density function of (Y, X, W). Equation (2.3) can be written as the operator equation

(2.4) $$T(h)(w) = qf_W(w),$$

where the operator T is defined by

$$T(h)(w) = \int_0^1 F_{YXW}[h(x), x, w]dx$$

for any function h for which the integral exists. Thus,

$$g = qT^{-1} f_W.$$

T and f_W are unknown but can be estimated consistently using standard methods. However, T^{-1} is a discontinuous operator (Horowitz and Lee 2007). Consequently, even if T were known, g could not be estimated consistently by replacing f_W with a consistent estimator. This is called the ill-posed inverse problem and is familiar in the literature on integral equations. See, for example, Groetsch (1984); Engl, Hanke, and Neubauer (1996); and Kress (1999). Because of the ill-posed inverse problem, the fastest possible rate of convergence of an estimator of g is typically much slower than the usual nonparametric rates. Depending on the details of the distribution of (Y, X, W), the rate may be slower than $O_p(n^{-\varepsilon})$ for any $\varepsilon > 0$ (Chen and Reiss 2007, Hall and Horowitz 2005). Because of the ill-posed inverse problem and consequent slow convergence of any estimator of g, a test based on comparing estimates of g and G will have low power.

The test developed here does not require nonparametric estimation of g and is not affected by the ill-posed inverse problem. Therefore, the "precision" of the test is greater than that of a nonparametric estimator of g. Let n denote the sample size used for testing. Under mild conditions, the test rejects H_0 with probability approaching 1 as $n \to \infty$ whenever $g(x) \neq G(x)$ on a set of non-zero probability. Moreover, like the test of Blundell and Horowitz (2007), the test developed here can detect a large class of structural functions g whose distance from the conditional quantile function G in a suitable metric is $O(n^{-1/2})$. In contrast, the rate of convergence in probability of a nonparametric estimator of g is always slower than $O_p(n^{-1/2})$.[1]

Throughout the remaining discussion, we use an extended version of (2.1) and (2.2) that allows g to be a function of a vector of endogenous explanatory variables, X, and a set of exogenous explanatory variables, Z. We write this model as

(2.4) $\quad Y = g(X, Z) + U; \; P(U \leq 0 \mid Z = z, W = w) = q$

for almost every $(z, w) \in \text{supp}(Z, W)$, where Y and U are random scalars, 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 whose support is contained in a compact set that we take to be $[0,1]^r$ ($r \geq 0$). The compactness assumption is not restrictive because it can be satisfied by carrying out monotone increasing transformations of any components of X, W, and Z whose supports are not compact. If $r = 0$, then Z is not included in (2.4). W is an instrument for X.

The inferential problem is to test the null hypothesis, H_0, that

(2.5) $\quad P(U \leq 0 \mid X = x, Z = z) = q$

[1] 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).

except, possibly, if (x,z) belongs to a set of probability 0. This is equivalent to testing $P[g(X,Z)=G(X,Z)]=1$ or $P[Y-G(X,Z)\le 0\,|\,Z=z,W=w]=q$. The alternative hypothesis, H_1, is that (2.5) does not hold on some set that has non-zero probability or, equivalently, that $P[g(X,Z)=G(X,Z)]<1$. The data, $\{Y_i,X_i,Z_i,W_i: i=1,...,n\}$, are a simple random sample of (Y,X,Z,W).

2.2 *The Test Statistic*

To form the test statistic, let f_{YXZW}, f_{XZW}, and f_{ZW}, respectively, denote the denote the probability density functions of (Y,X,Z,W), (X,Z,W) and (Z,W). Define

$$F_{YXZW}(y,x,z,w)=\int_{-\infty}^{y} f_{YXZW}(u,x,z,w)du .$$

Let $G(x,z)$ denote the q conditional quantile of Y: $G(x,z)=Q_q(Y\,|\,X=x,Z=z)$. Then under H_0,

$$\tilde{S}(z,w)\equiv\int_{[0,1]^p} F_{YXZW}[G(x,z),x,z,w]dx-qf_{ZW}(z,w)=0 \quad (2.6)$$

for almost every $(z,w)\in\operatorname{supp}(Z,W)$. H_1 is equivalent to the statement that (2.6) does not hold on a set $\mathcal{B}\subset[0,1]^{p+r}$ with non-zero Lebesgue measure. A test statistic can be based on a sample analog of $\int\tilde{S}(z,w)^2\,dzdw$, but the resulting rate of testing is slower than $n^{-1/2}$ due to the need to estimate f_{ZW} and F_{YXZW} nonparametrically. The rate $n^{-1/2}$ can be achieved by carrying out an additional smoothing step. To this end, for $\xi_1,\xi_2\in[0,1]^p$ and $\zeta_1,\zeta_2\in[0,1]^r$, let $\ell(\xi_1,\zeta_1;\xi_2,\zeta_2)$ denote the kernel of a nonsingular integral operator, L, from $L_2[0,1]^{p+r}$ to itself. That is, L is defined by

$$(L\psi)(\xi_2,\zeta_2)=\int_{[0,1]^{p+r}}\ell(\xi_1,\zeta_1;\xi_2,\zeta_2)\psi(\xi_1,\zeta_1)d\xi_1 d\zeta_1 \quad (2.7)$$

and is nonsingular, where ψ is a function in $L_2[0,1]^{p+r}$. Then H_0 is equivalent to

(2.8) $S(z,w)\equiv$

$$\int_{[0,1]^{2p+r}} F_{YXZW}[G(x,\zeta),x,\zeta,\eta]\ell(\zeta,\eta,z,w)dxd\zeta d\eta-q\int_{[0,1]^{p+r}} f_{ZW}(\zeta,\eta)\ell(\zeta,\eta,z,w)d\zeta d\eta=0$$

for almost every $(z,w)\in\operatorname{supp}(Z,W)$. H_1 is equivalent to the statement that (2.8) does not hold on a set $\mathcal{B}\subset[0,1]^{p+r}$ with non-zero probability. The test statistic is based on a sample analog of $\int S(z,w)^2\,dzdw$. Basing the test of H_0 on $S(z,w)$ avoids the ill-posed inverse problem because $S(z,w)$ does not depend on g.

To form a sample analog of $S(z,w)$, let $\hat{G}^{(-i)}(x,z)$ be an estimator of $G(x,z)$ based on all the data except observation i. This estimator is described in detail in the next paragraph. Let $I(\cdot)$ denote the indicator function. It follows from (2.8) that

(2.9) $S(z,w) \equiv$

$$\int_{[0,1]^{2p+r}} dxd\zeta d\eta \int_{-\infty}^{\infty} dyI[y \le G(x,\zeta)] f_{YXZW}(y,x,\zeta,\eta)]\ell(\zeta,\eta;z,w) - q\int_{[0,1]^{p+r}} f_{ZW}(\zeta,\eta)\ell(\zeta,\eta;z,w)d\zeta d\eta = 0$$

$$= E_{YXZW}\{I[Y \le G(X,Z)] - q\}\ell(Z,W;z,w)$$

The sample analog is of $S(z,w)$ is obtained from (2.9) by replacing G with the estimator $\hat{G}^{(-i)}$, the population expectation E_{YXZW} with the sample average, and multiplying the resulting expression by $n^{1/2}$ to obtain a random variable that has a non-degenerate limiting distribution. The resulting scaled sample analog is

(2.10) $$\hat{S}_n(z,w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le \hat{G}^{(-i)}(X_i,Z_i)] - q\}\ell(Z_i,W_i,z,w).$$

The test statistic is

$$\tau_n = \int_{[0,1]^{p+r}} \hat{S}_n(z,w)^2 dzdw.$$

Under H_0,

$$\int S(z,w)^2 dzdw = 0,$$

so τ_n differs from 0 only due to random sampling errors. Therefore, H_0 is rejected if τ_n is larger than can be explained by random sampling errors. A method for obtaining the critical value of τ_n is presented in Section 3.

The estimator $\hat{G}^{(-i)}$ is a kernel nonparametric quantile regression estimator based on a boundary kernel that overcomes edge effects (Gasser and Müller 1979; Gasser, Müller, and Mammitzsch 1985). A boundary kernel with bandwidth $h > 0$ is a function $K_h(\cdot,\cdot)$ with the property that for all $\xi \in [0,1]$ and some integer $s \ge 2$

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

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

$$(2.12)\quad \int_0^\infty u^j \bar{K}(u)du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \le j \le s-1. \end{cases}$$

If ξ is close to 0, we can set $K_h(u,\xi) = \bar{K}(-u/h)$. There are 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. Let ρ_q be the check function: $\rho_q(y) = y[q - I(y \le 0)]$. The estimator of G is

$$(2.13)\quad \hat{G}^{(-i)}(x,z) = \arg\inf_a \sum_{\substack{j=1 \\ j \ne i}}^{n} \rho_q(Y_i - a) K_{p,h}(x - X_j, x) K_{r,h}(z - Z_j, z).$$

The test statistic τ_n is obtained by substituting (2.13) into (2.10).

3. ASYMPTOTIC PROPERTIES

This section presents the asymptotic properties of the test of exogeneity based on τ_n and explains how to obtain the critical value of τ_n.

3.1 *Regularity Conditions*

This section states the assumptions that are used to obtain the asymptotic properties of τ_n. The following notation is used. For any real $a > 0$, define $[a]$ as the largest integer less than or equal to a. Define $U = Y - g(X,Z)$ and $V = Y - G(X,Z)$. Let $\|\cdot\|$ denote the Euclidean metric. For any vector $\boldsymbol{x} = (x_1,...,x_d)$, function $f(\boldsymbol{x})$, and vector of non-negative integers $\boldsymbol{k} = (k_1,...,k_d)$, define $|\boldsymbol{k}| = k_1 + ... + k_d$ and

$$D^{\boldsymbol{k}} f(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{k}|}}{\partial x_1^{k_1} ... \partial x_d^{k_d}} f(\boldsymbol{x}).$$

For a set $\mathcal{X} \subset \mathbb{R}^d$ and positive constants $a, M < \infty$, define $C_M^a(\mathcal{X})$ as the class of continuous functions $f : \mathcal{X} \to \mathbb{R}$ such that $\|f\|_a \leq M$, where

$$\|f\|_a = \max_{|\boldsymbol{k}| \leq [a]} \sup_{\boldsymbol{x} \in \mathcal{X}} |D^{\boldsymbol{k}} f(\boldsymbol{x})| + \max_{|\boldsymbol{k}| \leq [a]} \sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \frac{|D^{\boldsymbol{k}} f(\boldsymbol{x}) - D^{\boldsymbol{k}} f(\boldsymbol{x}')|}{\|\boldsymbol{x} - \boldsymbol{x}'\|^{a-[a]}}$$

and derivatives on the boundary of $\mathcal{X}$ are one sided. Let $f_{V|XZ}(v | x, z)$ denote the probability density function of V conditional on $(X, Z) = (x, z)$, and let f_{XZ} denote the probability density function of (X, Z) whenever these density functions exist.

We make the following assumptions.

Assumption 1: (i) The support of (X, Z, W) is $[0,1]^{2p+r}$, where $\dim(X) = \dim(W) = p$ and $\dim(Z) = r$. (ii) (Y, X, Z, W) has a probability density function f_{YXZW} with respect to Lebesgue measure. (iii) There is a finite constant C_f such that $|f_{YXZW}(y, x, z, w)| \leq C_f$ for all (y, x, z, w). Moreover, $\partial f_{YXZW}(y, x, z, w) / \partial y$ exists and is continuous and bounded for all (y, x, z, w). (iv) The data $\{Y_i, X_i, Z_i, W_i : i = 1, ..., n\}$ are an independent random sample of (Y, X, Z, W).

Assumption 2: (i) $P(U \leq 0 | Z = z, W = w) = q$ for almost every $(z, w) \in [0,1]^{p+r}$. (ii) There is a finite constant C_g such that $|g(x, z)| \leq C_g$ for all $(x, z) \in [0,1]^{p+r}$. (iii) Equation (2.4) has a solution $g(x, z)$ that is unique except, possibly, for (x, z) in a set of Lebesgue measure zero.

Assumption 3: (i) The probability density function $f_{V|XZ}(v | X = x, Z = z)$ exists for all v and $(x, z) \in \text{supp}(X, Z)$. Moreover, for all v in a neighborhood of zero and $(x, z) \in [0,1]^{p+r}$, $f_{V|XZ}(v | x, z) \geq \delta$ for some $\delta > 0$, and $\partial f_{V|XZ}(v | x, z) / \partial v$ exists and is continuous. (ii) $f_{XZ}(x, z) \geq C_{XZ}$ for all $(x, z) \in [0,1]^{p+r}$ and some constant $C_{XZ} > 0$. (iii) $G(x, z) \in C_{C_g}^s([0,1]^{p+r})$ with $s > 3(p + r)/2$ and C_g as in assumption 2. (iv) $f_{YXZW} \in C_{C_g}^s(\text{supp}(Y) \times [0,1]^{2p+r})$. (v) There are a neighborhood $\mathcal{N}_v$ of $v = 0$ and a constant C_f such that $f_{V|XZ} \in C_{C_f}^s(\mathcal{N}_v \times [0,1]^{p+r})$ and $f_{XZ}(x, z) \in C_{C_f}^s([0,1]^{p+r})$.

Assumption 4: (i) The kernel K_h satisfies (2.11) for s as in assumption 3. (ii) There is a constant $C_K < \infty$ such that $|K_h(u, \xi) - K_h(u', \xi| \leq C_K \|u - u'\| / h$ for all u, u', and $\xi \in [0,1]$. (iii) For each $\xi \in [0,1]$, $K_h(u, \xi)$, considered as a function of u, is supported on $[(\xi - h)/h, \xi / h] \cap \mathcal{K}$ for some compact interval $\mathcal{K}$ that is independent of ξ. (iv) $\sup\{|K_h(u, \xi)|: h > 0, \xi \in [0,1], u \in \mathcal{K}\} < \infty$. (v) The bandwidth h satisfies $h = C_h n^{-b}$ where $C_h > 0$ is a finite constant and $1/(2s) < b < 1/[3(p + r)]$.

Assumption 5: (i) The operator L defined in (2.7) is nonsingular. (ii) There is a constant $C_\ell < \infty$ such that

$$\sup_{(z,w,\zeta,\eta)\in[0,1]^{2(p+r)}} |\ell(z,w;\zeta,\eta)| \le C_\ell ,$$

$$\sup_{(z,w,\zeta,\eta)\in[0,1]^{2(p+r)}} |\partial\ell(z,w;\zeta,\eta)/\partial\zeta| \le C_\ell ,$$

and

$$\sup_{(z,w,\zeta,\eta)\in[0,1]^{2(p+r)}} |\partial\ell(z,w;\zeta,\eta)/\partial\eta| \le C_\ell .$$

Assumptions 1 and 2 specify the model and properties of the random variables under consideration. Assumption 2(iii) requires the structural function g to be identified. Assumption 3 establishes smoothness conditions. Because of the curse of dimensionality, the smoothness of G, $f_{V|XZ}$, and f_{XZ} must increase as $p+r$ increases. Assumption 4 establishes properties of the kernel function and requires the estimator of G to be undersmoothed. Undersmoothing prevents the asymptotic bias of $\hat{G}^{(-i)}$ from dominating the asymptotic distribution of τ_n. K_h must be a higher-order kernel if $p+r \ge 2$.

3.2 *Asymptotic Properties of the Test Statistic under* H_0

To obtain the asymptotic distribution of τ_n under H_0, let f_{YXZ} denote the probability density function of (Y,X,Z). Define

$$B_n(\zeta,\eta) = n^{-1/2}\sum_{i=1}^{n}\Big\{\{I[Y_i \le g(X_i,Z_i)]-q\}\ell(Z_i,W_i;\zeta,\eta)$$

$$-\{I[Y_i \le G(X_i,Z_i)]-q\}\left[\frac{\int_{[0,1]^p} f_{YXZW}[G(X_i,Z_i),X_i,Z_i,w)\ell(Z_i,w;\zeta,\eta)dw}{f_{YXZ}[G(X_i,Z_i),X_i,Z_i]}\right]\Big\}$$

and

$$R(\zeta_1,\eta_1;\zeta_2,\eta_2) = E[B_n(\zeta_1,\eta_1)B_n(\zeta_2,\eta_2)] .$$

Define the operator Ω on $L_2([0,1]^{p+r})$ by

$$(\Omega\phi)(\zeta_2,\eta_2) = \int_{[0,1]^{p+r}} R(\zeta_1,\eta_1;\zeta_2,\eta_2)\phi(\zeta_1,\eta_1)d\zeta_1 d\eta_1 . \tag{3.1}$$

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

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

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

Under H_0, $G = g$, so knowledge of or estimation of g is not needed to obtain the asymptotic distribution of τ_n under H_0. This observation is used in the next section to obtain the critical value of τ_n.

3.3 *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

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

uniformly over t. Define

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

---

[2] R is a bounded function under the assumptions of Section 3.1. Therefore, Ω is a compact, completely continuous operator with discrete eigenvalues.

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.

In applications, K_{ε} can be chosen informally by sorting the $\hat{\omega}_j$'s in decreasing order and plotting them as a function of j. They typically plot as random noise near $\hat{\omega}_j = 0$ when j is sufficiently large. One can choose K_{ε} to be a value of j that is near the lower end of the "random noise" range. The rejection probability of the τ_n test is not highly sensitive to K_{ε}, so it is not necessary to attempt precision in making the choice.

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

$$\lambda(X, Z; \zeta, \eta) = \int_{[0,1]^p} f_{YXZW}[G(X,Z), X, Z, w)] \ell(Z, w; \zeta, \eta) dw .$$

Because $G = g$ under H_0,

$$B_n(\zeta, \eta) = n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le G(X_i, Z_i)] - q\} \left\{ \ell(Z_i, W_i; \zeta, \eta) - \frac{\lambda(X_i, Z_i; \zeta, \eta)}{f_{YXZ}[G(X_i, Z_i), X_i, Z_i]} \right\} .$$

An estimator of $R(\zeta_1, \eta_1; \zeta_2, \eta_2)$ that is consistent under H_0 can be obtained by replacing unknown quantities with estimators on the right-hand side of

$$R(\zeta_1, \eta_1; \zeta_2, \eta_2) = E\{I[Y \le G(X,Z)] - q\}^2 \left\{ \ell(Z, W; \zeta_1, \eta_1) - \frac{\lambda(X, Z; \zeta_1, \eta_1)}{f_{YXZ}[G(X,Z), X, Z]} \right\}$$

$$\times \left\{ \ell(Z, W; \zeta_2, \eta_2) - \frac{\lambda(X, Z, \zeta_2, \eta_2)}{f_{YXZ}[G(X,Z), X, Z]} \right\} .$$

To do this, let $\hat{f}_{YXZW}$ and $\hat{f}_{YXZ}$, respectively, be kernel estimators of f_{YXZW} and f_{YXZ} with bandwidths that converge to 0 at the asymptotically optimal rates. As is well known, $\hat{f}_{YXZW}$ and $\hat{f}_{YXZ}$ are consistent uniformly over the ranges of their arguments. Define

$$\hat{\lambda}(X_i, Z_i; \zeta, \eta) = \int_{[0,1]^p} \hat{f}_{YXZW}[\hat{G}^{(-i)}(X_i, Z_i), X_i, Z_i, w)]\ell(Z_i, w; \zeta, \eta)dw; \quad i = 1,..., n$$

and

$$\hat{R}(\zeta_1, \eta_1; \zeta_2, \eta_2) = n^{-1}\sum_{i=1}^{n}\{I[Y_i \le \hat{G}^{(-i)}(X_i, Z_i)] - q\}^2 \left\{\ell(Z_i, W_i; \zeta_1, \eta_1) - \frac{\hat{\lambda}(X_i, Z_i; \zeta_1, \eta_1)}{\hat{f}_{YXZ}[\hat{G}^{(-i)}(X_i, Z_i), X_i, Z_i]}\right\}$$

$$\times \left\{\ell(Z_i, W_i; \zeta_2, \eta_2) - \frac{\hat{\lambda}(X_i, Z_i; \zeta_2, \eta_2)}{\hat{f}_{YXZ}[\hat{G}^{(-i)}(X_i, Z_i), X_i, Z_i]}\right\}.$$

Let $\hat{\Omega}$ be the operator defined by

$$(\hat{\Omega}\phi)(\zeta_2, \eta_2) = \int_{[0,1]^{p+r}} \hat{R}(\zeta_1, \eta_1; \zeta_2, \eta_2)\phi(\zeta_1, \eta_1)d\zeta_1 d\eta_1 ,$$

Denote the eigenvalues of $\hat{\Omega}$ by $\{\hat{\omega}_j : j = 1, 2, ...\}$ and order them so that $\hat{\omega}_1 \ge \hat{\omega}_2 \ge ... \ge 0$. The relation between the $\hat{\omega}_j$'s and ω_j's is given by the following theorem.

<u>Theorem 2</u>: Let assumptions 1-5 hold. Then $\hat{\omega}_j - \omega_j = o_p(1)$ 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 $\{\ell(Z_i, W_i; \zeta_1, \eta_1) - \hat{\lambda}(X_i, Z_i; \zeta_1, \eta_1) / \hat{f}_{YZX}[\hat{G}^{(-i)}(X_i, Z_i), X_i, Z_i]\}$, and let Υ denote the $n \times n$ diagonal matrix whose (i, i) element is $\{I[Y_i \le \hat{G}^{(-i)}(X_i, Z_i)] - q\}^2$. Then

$$\hat{R}(\zeta_1, \eta_1; \zeta_2, \eta_2) = n^{-1}\hat{F}(\zeta_1, \eta_1)'\Upsilon\hat{F}(\zeta_2, \eta_2).$$

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

$$\ell(Z, W; \zeta, \eta) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} d_{jk}\phi_j(\zeta, \eta)\phi_k(Z, W),$$

where

$$d_{jk} = \int_{[0,1]^{2(p+r)}} \ell(z, w; \zeta, \eta)\phi_j(\zeta, \eta)\phi_k(z, w)dwdzd\zeta d\eta ,$$

and

$$\hat{\lambda}(X, X; \zeta, \eta) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}\phi_j(\zeta, \eta)\phi_k(X, Z),$$

where

$$a_{jk} = \int_{[0,1]^{2(p+r)}} \hat{\lambda}(z, w; \zeta, \eta)\phi_j(\zeta, \eta)\phi_k(z, w)dwdzd\zeta d\eta .$$

Approximate $\ell(Z,W;\zeta,\eta)$ and $\hat{\lambda}(X,X;\zeta,\eta)$ by the finite sums

$$\Pi_{\ell}(Z,W;\zeta,\eta)=\sum_{j=1}^{L}\sum_{k=1}^{L}d_{jk}\phi_j(\zeta,\eta)\phi_k(Z,W)$$

and

$$\Pi_{\hat{\lambda}}(X,Z;\zeta,\eta)=\sum_{j=1}^{L}\sum_{k=1}^{L}a_{jk}\phi_j(\zeta,\eta)\phi_k(X,Z)$$

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

$$\Phi_{ij}=n^{-1/2}\sum_{k=1}^{L}\{d_{jk}\phi_k(Z_i,W_i)-a_{jk}\phi_k(X_i,Z_i)/\hat{f}_{YXZ}[\hat{G}^{(-i)}(X_i,Z_i),X_i,Z_i]\}.$$

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

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

In this section, it is assumed that H_0 is false. That is, $P[g(X,Z)=G(X,Z)]<1$. Define

(3.2) $$H(\zeta,\eta)=\int_{[0,1]^{2p+r}}\{F_{YXZW}[G(x,z),x,z,w]-F_{YXZW}[g(x,z),x,z,w]\}\ell(z,w;\zeta,\eta)dxdwdz\,.$$

Let z_α denote the $1-\alpha$ quantile of the asymptotic distribution of τ_n under sampling from the null-hypothesis model $Y=G(X,Z)+V, P(V\le 0\,|\,X,Z)=q$. The following theorem establishes consistency of the τ_n test against a fixed alternative hypothesis.

Theorem 3: Let assumptions 1-5 hold, and suppose that

$$\int_{[0,1]^{p+r}}H(\zeta,\eta)^2 d\zeta d\eta>0\,.$$

Then for any α such that $0<\alpha<1$,

$$\lim_{n\to\infty}\boldsymbol{P}(\tau_n>z_\alpha)=1. \quad \blacksquare$$

Because ℓ is the kernel of a nonsingular integral operator, 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.

3.5 *Asymptotic Distribution under Local Alternatives*

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

(3.3) $$P[Y\le G(X,Z)+n^{-1/2}\Delta(X,Z)\,|\,W=w,Z=z)=q$$

for almost every $(w,z) \in [0,1]^{p+r}$, where Δ is a bounded function on $[0,1]^{p+r}$. Under (3.3)

(3.4) $\quad g(x,z) = G(X,Z) + n^{-1/2}\Delta(x,z)$,

and

$$Y = g(X,Z) + U; \quad P(U \le 0 \mid Z = z, W = w) = q$$

for almost every $(w,z) \in [0,1]^{p+r}$.

Let Ω be the integral operator defined in (3.1), $\{\phi_j\}$ denote the orthornormal eigenfunctions of Ω, and $\{\omega_j\}$ denote the eigenvalues of Ω sorted so that $\omega_1 \ge \omega_2 \ge \ldots$ Let f_{UXZW} denote the probability density function of (U,X,Z,W). Define

$$\mu(\zeta,\eta) = -\int_{[0,1]^{2p+r}} f_{UXZW}(0,x,z,w)\Delta(x,z)\ell(z,w;\zeta,\eta)dxdzdw$$

and

(3.5) $\quad \mu_j = \int_{[0,1]^{p+r}} \mu(\zeta,\eta)\phi_j(\zeta,\eta)d\zeta d\eta$.

Let $\{\chi_1^2(\mu_j^2/\omega_j): j = 1,2,\ldots\}$ denote a sequence of independent random variables distributed as non-central chi-square with one degree of freedom and non-central parameters μ_j^2/ω_j.

The following theorem gives the asymptotic distribution of τ_n under the sequence of local alternatives (3.3)-(3.4).

<u>Theorem 4</u>: Let assumptions 1-5 hold. Under the sequence of local alternatives (3.3)-(3.4),

$$\tau_n \to^d \sum_{j-1}^{\infty} \omega_j \chi_{1j}^2(\mu_j^2/\omega_j). \quad \blacksquare$$

It follows from Theorems 2 and 4 that under (3.3)-(3.4),

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

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

$$\lim_{n\to\infty} P(\tau_n > z_\alpha) > \alpha$$

if $\mu_j^2 > 0$ for at least one j. In addition, for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(\tau_n > z_\alpha) > 1 - \varepsilon$$

if μ_j^2 is sufficiently large for at least one j.

3.6 *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 $O(n^{-1/2})$. This set contains deviations from H_0 that cannot be represented as sequences of local alternatives. Thus, the set is larger than the class of local alternatives against which the power of τ_n exceeds $1-\varepsilon$. The practical consequence of this result is to define a relatively large class of alternatives against which the τ_n test has high power in large samples.

The following additional notation is used. Let $\|\cdot\|$ denote the norm in $L_2[0,1]$. Define $H(\zeta,\eta)$ as in (3.2). Define the linear operator T by

$$(T\psi)(\zeta,\eta) = \int_{[0,1]^{2p+r}} f_{YXZW}[g(x,z),x,z,w]\ell(z,w;\zeta,\eta)\psi(x,z)dxdwdz$$

and the function

$$\pi(x,z) = g(x,z) - G(x,z).$$

For some finite $C > 0$, let $\mathcal{F}_{nC_g}$ be the class of functions $g(x,z) \in C^a_{C_g}([0,1]^{p+r})$ with $a > p+r$, $C_g < \infty$ satisfying:

(i) There is a function $G(x,z)$ such that $P[Y \le G(X,Z) \mid X = x, Z = z) = q$ for almost every $(x,z) \in [0,1]^{p+r}$.

(ii) Assumption 3 is satisfied with $V = Y - G(X,Z)$.

(iii) The density function f_{YZXW} satisfies Assumption 1.

(iv) The function g satisfies Assumption 2 with $U = Y - G(X,Z)$.

(v) $\|T\pi\| \ge n^{-1/2}C$

Condition (v) implies that $\mathcal{F}_{nC}$ contains alternative models g such that $\|g - G\| = O(n^{-1/2})$. In addition, condition (v) rules out differences between the structural functions under the null and alternative hypotheses, $\pi(x,z) = g(x,z) - G(x,z)$, that are linear combinations of eigenfunctions of T associated with eigenvalues of T that converge to zero too rapidly. Thus, the τ_n test has low power against deviations from H_0 that operate through eigenfunctions of T associated with eigenvalues that converge to zero very rapidly. Such deviations often correspond to highly oscillatory functions that have little relevance for economic applications.

The following theorem states the result of this section.

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

$$\lim_{n\to\infty} \inf_{g\in\mathcal{F}_{nC_g}} \boldsymbol{P}(\tau_n > z_\alpha) \geq 1-\delta$$

and

$$\lim_{n\to\infty} \inf_{g\in\mathcal{F}_{nC_g}} \boldsymbol{P}(\tau_n > \hat{z}_{\varepsilon\alpha}) \geq 1-2\delta . \quad \blacksquare$$

3.7 *Weight functions*

This section considers the choice of the weight function $\ell(z,w;\zeta,\eta)$. We show that setting $\ell(z,w;\zeta,\eta) = \ell_1(z,\zeta) f_{UXZW}(0,x,z,w)$ has certain power advantages over a weight function that does not depend on the distribution of (U,X,Z,W). The function ℓ_1 is assumed to be the kernel of a non-singular integral operator from $L_2([0,1]^r)$ to itself. Horowitz and Lee (2009) present a method for estimating $f_{UXZW}(0,x,z,w)$. Section 6.2 outlines the extension of Theorems 1-5 to the case of an estimated weight function.

To start, assume that $r = 0$, so Z is not in the model. Let τ_{nf} denote the τ_n statistic with weight function $f_{UXW}(0,x,w)$ and $\tau_{n\ell}$ denote the statistic with a fixed weight function $\ell(w,\eta)$ that does not depend on the distribution of (U,X,W). The arguments of Horowitz and Lee (2009) show that there are combinations of density functions f_{UXW} and local alternative models such that an α-level test based on $\tau_{n\ell}$ has local power that is arbitrarily close to α, whereas the asymptotic local power of an α-level test based on τ_{nf} is bounded away from and above α. In contrast, it is not possible for the asymptotic local power of the α-level τ_{nf} test to approach α while the asymptotic local power of the α-level $\tau_{n\ell}$ test remains bounded away from and above α.

Horowitz and Lee (2009) did not investigate the case of $r \geq 1$. The following theorem extends their result to this case.

Theorem 6: Let assumptions 1-5 hold. Let $\Delta(x,z)$ be the bounded function defined in (3.3)-(3.4). Fix the functions $\ell(z,w;\zeta,\eta)$ and $\ell_1(z,\zeta)$, and assume that these functions are bounded and that ℓ_1 is bounded away from 0. Define

$$\mu_\ell(\zeta,\eta) = \int f_{UXZW}(0,x,z,w)\Delta(x,z)\ell(z,w;\zeta,\eta)dxdzdw$$

and

$$\mu_f(\zeta,\eta) = \int f_{UXZW}(0,x,z,w)\Delta(x,z)\ell_1(z,\zeta)\, f_{UXZW}(0,\eta,z,w)dxdzdw .$$

Then

(a) For any $\varepsilon > 0$, there are functions $\Delta(x,z)$ and f_{UXZW} such that $\|\mu_\ell\|^2 / \sum_{j=1}^{\infty}\omega_j < \varepsilon$ and $\|\mu_f\|^2 / \sum_{j=1}^{\infty}\omega_j \ge D_1^2$ for some $D_1^2 > 0$.

(b) There is a constant $D > 0$ such that $\|\mu_\ell\|^2 \le D\|\mu_f\|^2$. ■

Theorem 6(a) implies that there are combinations of density functions f_{UXW} and local alternative models such that an α-level test based on $\tau_{n\ell}$ has local power that is arbitrarily close to α, whereas the asymptotic local power of an α-level test based on τ_{nf} is bounded away from and above α. Theorem 6(b) implies that it is not possible for the asymptotic local power of the α-level τ_{nf} test to approach α while the asymptotic local power of the α-level τ_{nf} test remains bounded away from and above α.

Theorem 6 does not imply that the power of τ_{nf} always exceeds that of $\tau_{n\ell}$. Moreover, in finite samples, random sampling errors in an estimate of f_{UXZW} can reduce the power of τ_{nf} and increase the difference between the true and nominal probabilities of rejecting a correct H_0. Consequently, a weight function that does not depend on the sample may be attractive in applications. Section 4 provides illustrations of the finite-sample performances of τ_{nf} and $\tau_{n\ell}$ with two weight functions that do not depend on the sample.

4. 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,U) were generated by

$$W = \Phi(\zeta),$$

$$X = \Phi\left(\rho_1\zeta + \sqrt{1-\rho_1^2}\,\xi\right),$$

and

$$U = \rho_2\xi + \sqrt{1-\rho_2^2}\,\nu ,$$

where Φ is the $N(0,1)$ distribution function; ζ, ξ, and ν are independent random variables with $N(0,1)$ distributions; and ρ_1 and ρ_2 ($0 \le \rho_1, \rho_2 \le 1$) are constant parameters whose values vary among experiments. The parameter ρ_1 determines the strength of the instrument W, and ρ_2 determines the strength of the correlation between U and X. H_0 is true if $\rho_2 = 0$ and false otherwise. Realizations of Y were generated from

(4.1) $\quad Y = \theta_0 + \theta_1 X + \sigma_U U$,

where $\theta_0 = 0$, $\theta_1 = 0.5$, and $\sigma_U = 0.1$. Experiments were carried out with $\rho_1 = 0.35$ or 0.7, and $\rho_2 = 0,\ 0.1,\ 0.2$, or 0.3. The instrument is stronger when $\rho_1 = 0.7$ than when $\rho_1 = 0.35$, and the correlation between X and U increases as ρ_2 increases. The sample size was $n = 750, 1000$, or 2000, depending on the experiment, and the nominal probability of rejecting a correct H_0 was 0.05. There were 2000 Monte Carlo replications per experiment.

The kernel function $K(v) = (15/16)(1-v^2)^2 I(|v| \le 1)$ was used to compute $\hat{G}^{(-i)}$ in τ_n and $\hat{f}_{YXW}$ in the estimated critical value of τ_n and in the data-dependent weight function. The rule-of-thumb bandwidth of Yu and Jones (1998) was used for $\hat{G}^{(-i)}$ and $\hat{f}_{YXW}$ in the critical value of τ_n. Four different weight functions $\ell(w,\eta)$ were used in τ_n. One is the data-dependent estimated probability density function $\hat{f}_{YXW}[\hat{g}(\eta),\eta,w]$ with $\hat{g}$ computed using the method of Horowitz and Lee (2009). The bandwidths for $\hat{f}_{YXW}$ in the Horowitz-Lee estimator were $h_X = h_Y = 0.01$ for the X and Y directions and $h_W = 0.3$ for the W direction. The other weight functions are not data dependent. The second weight function is the infeasible true probability density function $f_{YXW}[g(\eta),\eta,w]$. The third and fourth weight functions are $\ell(w,\eta) = I(w \le \eta)$ and $\ell(w,\eta) = \exp(w\eta)$, respectively. The third weight function was used by Song (2010) and Stute and Zhu (1998). The fourth was proposed by Bierens (1990). The second weight function is not feasible in applications but provides an indication of the reduction in finite-sample performance due to random sampling errors in estimating the weight function.

The results of the experiments are shown in Table 1 for $\rho_1 = 0.35$ and Table 2 for $\rho_1 = 0.7$. In the tables, τ_{nD}, τ_{nD}^*, τ_{nI}, and τ_{nB}, respectively, denote the τ_n tests with the data-dependent weight function, the infeasible weight function, the Song (2010) weight function, and the Bierens (1990) weight function. In what follows, the difference between the empirical and nominal probabilities of rejecting a correct H_0 is called the error in the rejection probability or ERP. The performance of the τ_{nB} test is poor. It has a large ERP when $n < 2000$ and low power. The τ_{nI} test has the best performance over all experiments. Its ERP is low. Its power is higher than the that of the τ_{nB} test and only slightly lower than the power of the infeasible τ_{nD}^* test in experiments in which the τ_{nD}^* test has a low ERP. The τ_{nD} test has a relatively high ERP if the instrument is weak ($\rho_1 = 0.35$) or n is small. The power of the τ_{nD} test is lower than that of the τ_{nD}^* test. The relatively poor performance of τ_{nD} compared to τ_{nD}^* is a consequence of random sampling errors in estimating $f_{YXW}[g(x),\eta,w]$ in τ_{nD}. In summary, the τ_{nI} test performs particularly well in the Monte Carlo experiments. It has good power and a low ERP, even with moderate sample sizes.

5. CONCLUSIONS

Endogeneity of explanatory variables is an important problem in applied econometrics. Erroneously assuming that explanatory variables are exogenous can lead to highly misleading results. This paper has described a test for exogeneity in nonparametric quantile regressions. The test does not use a parametric model, thereby avoiding the possibility of obtaining misleading results due to misspecification of the model. The test also avoids the slow rate of convergence and potentially low power associated with the ill-posed inverse problem of nonparametric instrumental variables estimation of either mean- or quantile-regression models. The new test has non-trivial power against alternative hypotheses whose "distance" from the null hypothesis of exogeneity is $O(n^{-1/2})$, which is the same as the distance possible with tests based on parametric models. The results of Monte Carlo experiments have illustrated the finite-sample performance of the test.

6. APPENDIX: PROOFS OF THEOREMS AND EXTENSION TO AN ESTIMATED WEIGHT FUNCTION

6.1 *Proofs of Theorems 1*-6

Assumptions 1-5 hold throughout this section. To minimize the complexity of the proofs without losing any important elements, assume that $p=1$ and $r=0$. The proofs with $p>1$ and $r>0$ are identical after replacing quantities for $p=1$ and $r=0$ with analogous quantities for the more general case. Let f_{YXW} and f_{YX}, respectively, denote the probability density functions of (Y,X,W) and (Y,X). With $p=1$ and $r=0$, (2.10) becomes

$$(6.1)\qquad \hat{S}_n(w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le \hat{G}^{(-i)}(X_i)] - q\}\ell(W_i,w),$$

$\lambda(X,X;\zeta,\eta)$ becomes

$$\lambda(X;\eta) = \int_0^1 f_{YXW}[G(X),X,w]\ell(w,\eta)dw,$$

and the test statistic is

$$\tau_n = \int_0^1 \hat{S}_n(w)^2 \, dw .$$

Define

$$S_{n1}(w) = n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le g(X_i)] - q\} \ell(W_i, w) ,$$

$$S_{n2}(w) = n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le G(X_i)] - I[Y_i \le g(X_i)]\} \ell(W_i, w) ,$$

and

$$S_{n3}(w) = n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le \hat{G}^{(-i)}(X_i)] - I[Y_i \le G(X_i)]\} \ell(W_i, w) .$$

Then

$$\hat{S}_n(w) = \sum_{j=1}^{3} S_{nj}(w) .$$

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

$$S_{n3}(w) = -n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le G(X_i)] - q\} \frac{\lambda(X_i, w)}{f_{YX}[G(X_i), X_i]} + o_p(1)$$

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

<u>Proof</u>: Write $S_{n3}(w) = S_{n31}(w) + S_{n32}(w)$, where

$$S_{n31}(w) = S_{n3}(w) - E_i[S_{n3}(w)] ,$$

$$S_{n32}(w) = E_i[S_{n3}(w)] ,$$

and E_i denotes the expectation over random variables indexed by i. It follows from Theorem 2.1 of van der Vaart and Wellner (2007) and the consistency and asymptotic Gaussianity of nonparametric quantile regression estimators that

$$\sup_{w \in [0,1]} | S_{n31}(w) | = o_p(1) . \tag{6.2}$$

Therefore, the lemma follows if

$$S_{n32}(w) = -n^{-1/2} \sum_{i=1}^{n} \{I[Y_i \le G(X_i)] - q\} \frac{\lambda(X_i, w)}{f_{YX}[G(X_i), X_i]} + o_p(1) \tag{6.3}$$

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

To prove (6.3), observe that

$$S_{n32}(\nu) = n^{-1/2} \sum_{i=1}^{n} \int \{F_{YXW}[\hat{G}^{(-i)}(x), x, w] - F_{YXW}[G(x), x, w]\} \ell(w, \nu) \, dx dw . \tag{6.4}$$

A Taylor series expansion yields

$$\int \{F_{YXW}[\hat{G}^{(-i)}(x),x,w]-F_{YXW}[G(x),x,w]\}\ell(w,\nu)dxdw$$

$$=\int f_{YXW}[G(x),x,w][\hat{G}^{(-i)}(x)-G(x)]\ell(w,\nu)dxdw+O\left(\left\|\hat{G}^{(-i)}-G\right\|_{\infty}^{2}\right).$$

Therefore,

$$S_{n32}(\nu)=^{a.s}-n^{-1/2}\int f_{YXW}[G(x),x,w][\hat{G}^{(-i)}(x)-G(x)]\ell(w,\nu)dxdw+o(1) \tag{6.5}$$

Calculations like those in Kong, Linton, and Xia (2010) show that

$$\hat{G}^{(-i)}(x)-G(x)=\frac{1}{f_{YX}[G(x),x]}\frac{1}{nh}\sum_{\substack{j=1\\ j\neq i}}^{n}\{q-I[Y_j\le G(X_j)]\}K\left(\frac{X_j-x}{h}\right)+R_n(x),$$

where

$$\sup_{x\in[0,1]}|R_n(x)|=^{a.s.}O\left[\left(\frac{\log n}{nh}\right)^{3/4}+h^s\right].$$

Therefore, standard calculations for kernel estimators yield

$$\int \{F_{YXW}[\hat{G}^{(-i)}(x),x,w]-F_{YXW}[G(x),x,w]\}\ell(w,\nu)dx \tag{6.6}$$

$$=^{a.s}n^{-1}\sum_{j=1}^{n}\{q-I[Y_j\le G(X_j)]\}\frac{\lambda(X_j,\nu)}{f_{YX}[G(X_j),X_j]}+O\left[\left(\frac{\log n}{nh}\right)^{3/4}+h^s\right].$$

The lemma follows by substituting (6.6) into (6.4). Q.E.D.

Proof of Theorem 1: Under H_0, $S_{n2}(w)=0$ and $g=G$. Therefore, it follows from Lemma 1 that

$$S_n(\eta)=S_{n1}(\eta)+S_{n3}(\eta)$$

$$=B_n(\eta)+o_p(1)$$

uniformly over $\eta\in[0,1]$, where

$$B_n(\eta)=n^{-1/2}\sum_{i=1}^{n}\{I[Y_i\le G(X_i)]-q\}\left\{\ell(W_i;\eta)-\frac{\lambda(X_i;\eta)}{f_{YX}[G(X_i),X_i]}\right\}.$$

Therefore,

$$\tau_n\to^{d}\int B_n^2(\eta)d\eta\ .$$

But

$$B_n(\eta)=\sum_{j=1}^{\infty} b_j\phi_j(\eta)$$

where the ϕ_j 's are the eigenfunctions of the operator Ω defined in (3.1) and

$$b_j=\int_0^1 B_n(\eta)\phi_j(\eta)d\eta .$$

It follows that

$$\tau_n \to^d \sum_{j=1}^{\infty} b_j^2=\sum_{j=1}^{n}\omega_j\left(\frac{b_j}{\omega_j^{1/2}}\right)^2 .$$

The b_j 's are independently distributed as $N(0,\omega_j)$. Therefore, the random variables $(b_j/\omega_j)^2$ are independently distributed as chi-square with one degree of freedom. Q.E.D.

Proof of Theorem 2: Let $\|\cdot\|_2$ denote the L_2 norm and $\|\cdot\|_{op}$ denote the operator norm

$$\|A\|_{op}=\sup_{\|u\|_2}\|Au\|_2 ,$$

where A is an operator on $L_2[0,1]$. By Theorem 5.1a of Bhatia, Davis, and McIntosh (1983), it suffices to prove that $\|\hat{\Omega}-\Omega\|_{op}\to^p 0$ as $n\to\infty$. An application of the Cauchy-Schwarz inequality shows that

$$\|\hat{\Omega}-\Omega\|_{op}\le\int_{[0,1]^2}[\hat{R}(\eta_1;\eta_2)-R(\eta_1;\eta_2)]^2 d\eta_1 d\eta_2 .$$

It follows from uniform consistency of $\hat{G}^{(-i)}$ for G, $\hat{f}_{YXZW}$ for f_{YXZW}, and $\hat{f}_{YXW}$ for f_{YXW} that

$$\hat{R}(\eta_1,\eta_2)=\tilde{R}(\eta_1,\eta_2)+o_p(1)$$

uniformly over $\eta_1,\eta_2\in[0,1]^2$, where

$$\tilde{R}(\eta_1;\eta_2)$$

$$\equiv n^{-1}\sum_{i=1}^{n}\{I[Y_i\le\hat{G}^{(-i)}(X_i)]-q\}^2\left\{\ell(W_i;\eta_1)-\frac{\lambda(X_i,\eta_1)}{f_{YX}[G(X_i),X_i]}\right\}\left\{\ell(W_i;\eta_2)-\frac{\lambda(X_i,\eta_2)}{f_{YX}[G(X_i),X_i]}\right\}.$$

Arguments like those used to prove lemma 1 show that $\tilde{R}_n(\eta_1,\eta_2)=R(\eta_1,\eta_2)+o_p(1)$ for each η_1,η_2, so $\hat{R}_n(\eta_1,\eta_2)=R(\eta_1,\eta_2)+o_p(1)$ as $n\to\infty$ for each η_1,η_2. Therefore,

$$\int_{[0,1]^2}[\hat{R}(\eta_1;\eta_2)-R(\eta_1;\eta_2)]^2 d\eta_1 d\eta_2=o_p(1)$$

by the dominated convergence theorem. Q.E.D.

Proof of Theorem 3: Let $\int_0^1 H(\eta)^2 d\eta>0$. It suffices to show that

$$\operatorname*{plim}_{n\to\infty} n^{-1}\tau_n > 0 .$$

As $n \to \infty$,

$$n^{-1/2} S_{n1}(\eta) \to^{a.s.} 0$$

and

$$n^{-1/2} S_{n2}(\eta) \to^{a.s} \int_0^1 \{F_{YXW}[G(x),x] - F_{YXW}[g(x),x]\}\ell(w,\eta)dxdw = H(\eta) .$$

by the strong law of large numbers (SLLN). In addition

$$n^{-1/2} S_{n3}(\eta) \to^{p} 0$$

by lemma 1 and the SLLN. Therefore,

$$n^{-1}\tau_n \to^{p} \int_0^1 H(\eta)^2 d\eta > 0 .$$

Q.E.D.

Proof of Theorem 4: By lemma 1

$$\hat{S}_n(\eta) = S_{n1}(\eta) + S_{n2}(\eta) + S_{n3}(\eta)$$

$$= B_n(\eta) + S_{n2}(\eta) + o_p(1).$$

Some algebra shows that $E[S_{n2}(\eta)] = \mu(\eta)$ and $Var[S_{n2}(\eta)] = O(n^{-1/2})$. Therefore, $S_{n2}(\eta) \to^{p} \mu(\eta)$,

$$\hat{S}_n(\eta) = B_n(\eta) + \mu(\eta) + o_p(1) ,$$

and

$$\tau_n \to^{d} \int_0^1 [B_n(\eta) + \mu(\eta)]^2 d\eta .$$

But

$$B_n(\eta) = \sum_{j=1}^{\infty} b_j \phi_j(\eta)$$

and

$$\mu(\eta) = \sum_{j=1}^{\infty} \mu_j \phi_j(\eta) ,$$

where the ϕ_j 's are the eigenfunctions of the operator Ω defined in (3.1), the μ_j 's are as defined in (3.5), and

$$b_j = \int_0^1 B_n(\eta)\phi_j(\eta)d\eta .$$

It follows that

$$\tau_n \to^d \sum_{j=1}^{\infty}(b_j+\mu_j)^2 = \sum_{j=1}^{n}\omega_j\left(\frac{b_j}{\omega_j^{1/2}}+\frac{\mu_j}{\omega_j^{1/2}}\right)^2 .$$

The b_j 's are independently distributed as $N(0,\omega_j)$, and the μ_j 's are non-stochastic. Therefore, the random variables $\left(\frac{b_j}{\omega_j^{1/2}}+\frac{\mu_j}{\omega_j^{1/2}}\right)^2$ are independently distributed as non-central chi-square with one degree of freedom and non-central parameter μ_j^2/ω_j. Q.E.D.

Proof of Theorem 5: Define

$$S_{n2}^*(\eta) = E\{I[Y \le G(X) - I[Y \le g(X)]\}\ell(W,\eta),$$

$$D_n(\eta) = S_{n3}(\eta) + n^{1/2}S_{n2}^*(\eta),$$

and

$$\tilde{S}_n(\eta) = S_n(\eta) - D_n(\eta).$$

Then

$$\tilde{S}_n(\eta) = S_{n1}(\eta) + [S_{n2}(\eta) - ES_{n2}(\eta)]$$

$$= n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le G(X_i)]\ell(W_i,\eta) - EI[Y \le G(X)]\ell(W,\eta)\}$$

$$- n^{-1/2}\sum_{i=1}^{n} q[\ell(W_i,\eta) - E\ell(W,\eta)].$$

It follows from lemma (2.13) of Pakes and Pollard (1989) and Theorem 7.21 of Pollard (1984) that $\tilde{S}_n(\eta)$ and $\|\tilde{S}_n\|$ are bounded in probability uniformly over $\eta \in [0,1]$.

Note that $\tau_n = \|S_n\|^2$. Use the inequality $a^2 \ge 0.5b^2 - (b-a)^2$ with $a = S_n$ and $b = D_n$ to obtain

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

Because $\|\tilde{S}_n\| = O_p(1)$, for each $\varepsilon > 0$ there is $M_\varepsilon < \infty$ such that for all $M > M_\varepsilon$,

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

$$\le P\left(0.5\|D_n\|^2 < z_\alpha + \le M\right) + \varepsilon.$$

Therefore,

(6.7) $\quad P(\tau_n > z_\alpha) \ge P\left(0.5\|D_n\|^2 > z_\alpha + M\right)$.

Now $D_n = S_{n3} + n^{1/2} S_{n2}^*$, and $S_{n2}^* = T\pi + o_p(n^{-1/2})$. By $a^2 \ge 0.5b^2 - (b-a)^2$ with $a = D_n$ and $b = n^{1/2} S_{n2}^*$,

$$\|D_n\|^2 \ge 0.5n\|S_{n2}^*\|^2 - \|S_{n3}\|^2$$
$$= 0.5n\|T\pi\|^2 - \|S_{n3}\|^2 + o_p(1).$$

But $\|S_{n3}\|^2 = O_p(1)$ by lemma 1. Therefore,

(6.8) $\quad \|D_n\|^2 \ge 0.5n\|T\pi\|^2 + O_p(1)$.

Substituting (6.8) into (6.7) yields

$$P(\tau_n > z_\alpha) \ge P\left(0.25n\|T\pi\|^2 + \xi_n > z_\alpha + M\right)$$

for some random variable $\xi_n = O_p(1)$. The theorem follows by letting C in the definition of $\mathcal{F}_{nC}$ be sufficiently large. Q.E.D.

Proof of Theorem 6:

Part (a): We construct an example in which $\|\mu_\ell\|^2 < \varepsilon$ and $\|\mu_f\|^2 = 1$. To simplify the discussion, assume that G is known and does not have to be estimated, and set $p = r = 1$. Define

$$\tilde{B}_{nf}(\zeta,\eta) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le g(X_i, Z_i)] - q\}\ell_1(Z_i,\zeta)\, f_{UXZW}(0,\eta,Z_i,W_i),$$

$$\tilde{B}_{n\ell}(\zeta,\eta) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le g(X_i, Z_i)] - q\}\ell(Z_i,W_i;\zeta,\eta),$$

$\tilde{R}_{nf}(\zeta_1,\eta_1;\zeta_2,\eta_2) = E[\tilde{B}_{nf}(\zeta_1,\eta_1)\tilde{B}_{nf}(\zeta_2,\eta_2)]$, and $\tilde{R}_{n\ell}(\zeta_1,\eta_1;\zeta_2,\eta_2) = E[\tilde{B}_{n\ell}(\zeta_1,\eta_1)\tilde{B}_{n\ell}(\zeta_2,\eta_2)]$. Also, define the operators $\tilde{\Omega}_f$ and $\tilde{\Omega}_\ell$ on $L_2([0,1]^2)$ by

$$(\tilde{\Omega}_f \vartheta)(\zeta_2, \eta_2) = \int_{[0,1]^2} \tilde{R}_f(\zeta_1, \eta_1; \zeta_2, \eta_2) \vartheta(\zeta_1, \eta_1) d\zeta_1 d\eta_1$$

and

$$(\tilde{\Omega}_\ell \vartheta)(\zeta_2, \eta_2) = \int_{[0,1]^2} \tilde{R}_\ell(\zeta_1, \eta_1; \zeta_2, \eta_2) \vartheta(\zeta_1, \eta_1) d\zeta_1 d\eta_1 .$$

Let $\{(\tilde{\omega}_{jf}, \tilde{\psi}_{jf});\ j = 1, 2, ...\}$ and $\{(\tilde{\omega}_{j\ell}, \tilde{\psi}_{j\ell});\ j = 1, 2, ...\}$ denote the eigenvalues and eigenvectors of $\tilde{\Omega}_f$ and $\tilde{\Omega}_\ell$, respectively, sorted in decreasing order of the eigenvalues. Define

$$\tilde{\mu}_f(\zeta, \eta) = -\int_{[0,1]^3} f_{UXZW}(0, x, z, w) \Delta(x, z) \ell_1(z, \zeta) f_{UXZW}(0, \eta, z, w) dxdzdw ,$$

$$\tilde{\mu}_\ell(\zeta, \eta) = -\int_{[0,1]^3} f_{UXZW}(0, x, z, w) \Delta(x, z) \ell(z, w; \zeta, \eta) dxdzdw ,$$

$$\tilde{\mu}_{jf} = \int_{[0,1]^2} \tilde{\mu}_f(\zeta, \eta) \tilde{\psi}_{jf}(\zeta, \eta) d\zeta d\eta ,$$

and

$$\tilde{\mu}_{j\ell} = \int_{[0,1]^2} \tilde{\mu}_\ell(\zeta, \eta) \tilde{\psi}_{j\ell}(\zeta, \eta) d\zeta d\eta .$$

Arguments identical to those used to prove Theorem 4 but with a known G show that under the sequence of local alternative hypotheses (3.3)-(3.4),

$$\tau_{nf} \to^d \sum_{j-1}^{\infty} \tilde{\omega}_{jf} \chi^2_{1j}(\tilde{\mu}^2_{jf} / \tilde{\omega}_{jf})$$

and

$$\tau_{n\ell} \to^d \sum_{j-1}^{\infty} \tilde{\omega}_{j\ell} \chi^2_{1j}(\tilde{\mu}^2_{j\ell} / \tilde{\omega}_{j\ell})$$

as $n \to \infty$.

To establish part (a), it suffices to show that for any fixed function ℓ, f_{UXZW} and Δ can be chosen so that $\left\|\tilde{\mu}^2_f\right\|^2 / \sum_{j=1}^{\infty} \tilde{\omega}_{jf}$ is bounded away from 0 and $\left\|\tilde{\mu}^2_\ell\right\|^2 / \sum_{j=1}^{\infty} \tilde{\omega}_{j\ell}$ is arbitrarily close to 0.

To do this, assume that Z is independent of (U, X, W) so that

$$f_{UZXW}(0, x, z, w) = f_Z(z) f_{UXW}(0, X, W) ,$$

where f_Z and f_{UXW}, respectively, are the probability density functions of Z and (U, X, W). For $v \in [0,1]$, define $\phi_1(v) = 1$ and $\phi_{j+1}(v) = 2^{-1/2} \cos(j\pi v)$ for $j \geq 1$. Define

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

Let

$$f_{UXW}(0,x,w) = 1 + \sum_{j=1}^{\infty} \lambda_j^{1/2} \phi_{j+1}(x)\phi_{j+1}(w).$$

Then

$$\tilde{R}_f(\zeta_1,\eta_1;\zeta_2,\eta_2) = q(1-q)E_Z[\ell_1(Z,\zeta_1)\ell_1(Z,\zeta_2) f_Z(Z)^2]E_W[f_{UXW}(0,\eta_1,W) f_{UXW}(0,\eta_2,W]$$

$$= q(1-q)Q(\zeta_1,\zeta_2)\left[1 + \sum_{j=1}^{\infty} \lambda_j \phi_{j+1}(\eta_1)\phi_{j+1}(\eta_2)\right],$$

where

$$Q(\zeta_1,\zeta_2) = E_Z[\ell_1(Z,\zeta_1)\ell_2(Z,\zeta_2) f_Z(Z)^2].$$

Let $\{\nu_k : k = 1,2,...\}$ denote the eigenvalues of the integral operator whose kernel is $Q(\zeta_1,\zeta_2)$. Then the eigenvalues of $\tilde{\Omega}_f$ are $\{\lambda_j \nu_k : j,k = 1,2,...\}$. Let

$$\Delta(x,z) = D_0 \Delta_Z(z)\phi_m(x),$$

for some $m \ge 1$, where $D_0 > 0$ is a constant and $\Delta_Z \in L_2([0,1])$ is a bounded function. Then

$$\tilde{\mu}_f(\zeta,\eta) = -D_0\phi_m(\eta)\int_0^1 \ell_1(z,\zeta) f_Z(z)^2 \Delta_Z(z)dz,$$

and

$$\|\tilde{\mu}_f\|^2 = D_0^2 \int_0^1 \left[\int_0^1 \ell_1(z,\zeta) f_Z(z)^2 \Delta_Z(z)dz\right]^2 d\zeta$$

$$\equiv D_1^2.$$

Moreover, $D_1^2 > 0$ for any m because ℓ_1 is the kernel of a non-singular integral operator.

We now show that m can be chosen so that $\|\mu_\ell\|^2$ is arbitrarily close to 0. To do this, observe that $\ell(z,w;\zeta,\eta)$ has the Fourier representation

$$\ell(z,w;\zeta,\eta) = \sum_{j,k,s,t=1}^{\infty} h_{jkst}\phi_j(z)\phi_k(w)\phi_s(\zeta)\phi_t(\eta),$$

where $\{h_{jkst} : j,k,s,t = 1,2,...\}$ are constants. Then

$$\tilde{\mu}_\ell(\zeta,\eta) = -D_0 \sum_{j,s,t=1}^{\infty} b_j h_{jmst}\phi_s(\zeta)\phi_t(\eta),$$

where

$$b_j = \int_0^1 f_Z(z)\Delta_Z(z)\phi_j(z)dz.$$

The b_j 's are Fourier coefficients of $f_Z(z)\Delta_Z(z)$, so $\sum_{j=1}^{\infty} b_j^2 = c_b$ for some $c_b < \infty$. Therefore, by the Cauchy-Schwarz inequality

$$\|\mu_\ell\|^2 = D_0^2 \sum_{s,t=1}^{\infty} \left(\sum_{j=1}^{\infty} b_j h_{jmst} \right)^2$$

$$\leq c_b D_0^2 \sum_{j,s,t=1}^{\infty} h_{jmst}^2 .$$

Because ℓ is bounded, m can be chosen so that

$$\sum_{j,s,t=1}^{\infty} h_{jmst}^2 < \varepsilon / (c_b D_0^2)$$

for any $\varepsilon > 0$. With this m, $\|\mu_\ell\|^2 < \varepsilon$, which establishes part (a).

Part (b): We have

$$\tilde{\mu}_\ell(\zeta,\eta) = -\int_{[0,1]^3} f_{UXZW}(0,x,z,w)\Delta(x,z)\ell(z,w;\zeta,\eta)dxdzdw .$$

By the Cauchy-Schwarz inequality,

$$\|\tilde{\mu}_\ell\|^2 \leq \int_{[0,1]^2} \left\{ \int_{[0,1]^2} \left[\int_{[0,1]} f_{UXZW}(0,x,z,w)\Delta(x,z)dx \right]^2 dzdw \times \left[\int_{[0,1]^2} \ell(z,w;\zeta,\eta)^2 dzdw \right] \right\} d\zeta d\eta$$

$$\leq C_\ell \int_{[0,1]^2} \left[\int_{[0,1]} f_{UXZW}(0,x,z,w)\Delta(x,z)dx \right]^2 dzdw \qquad (6.9)$$

for some constant $C_\ell < \infty$. Under assumption 2(ii), $\Delta(x,z)$ is bounded from below, say by $c_\Delta > -\infty$, so it can be assumed without loss of generality that $\Delta(x,z) \geq 0$ for all $(x,z) \in [0,1]^{p+r}$. (If $c_\Delta < 0$, replace $\Delta(x,z)$ by $\Delta(x,z) - c_\Delta$ and $G(x,z)$ by $G(x,z) + c_\Delta$. This is a normalization that has no effect on model (3.4) because G is nonparametric.) By the boundedness of $\Delta(x,z)$ from above, and of $\ell_1(z,\zeta)$ from below,

$$\int_{[0,1]^2}\left[\int_{[0,1]} f_{UXZW}(0,x,z,w)\Delta(x,z)dx\right]^2 dzdw$$

$$= \int_{[0,1]^4} f_{UXZW}(0,x,z,w)\Delta(x,z) f_{UXZW}(0,\eta,z,w)\Delta(\eta,z)dxd\eta dzdw$$

$$= \int_{[0,1]^5} f_{UXZW}(0,x,z,w)\Delta(x,z) f_{UXZW}(0,\eta,z,w)\Delta(\eta,z)dxd\eta dzdwd\zeta$$

$$\le C_1\int_{[0,1]^5} f_{UXZW}(0,x,z,w)\Delta(x,z) f_{UXZW}(0,\eta,z,w)\ell_1(\zeta,z)dxd\eta dzdwd\zeta$$

$$= C_1\int_{[0,1]^2} |\tilde{\mu}_f(\zeta,\eta)|\, d\zeta\, d\eta$$

$$\le C_1\left\|\mu_f\right\|^2 \tag{6.10}$$

for some finite constant $C_1 < \infty$, where the last line follows from the Cauchy-Schwarz inequality. Theorem 6(b) follows from substituting (6.10) into (6.9). Q.E.D.

6.2 *Extension of Theorems 1-5 to the case of an estimated weight function*

Let $\hat{\ell}(W,\eta)$ be an estimator of the weight function $\hat{\ell}(W,\eta)$. The test statistic with the estimated weight function, $p=1$, and $r=0$ is

$$\tau_n = \int_0^1 \breve{S}_n(w)^2 dw,$$

where

$$\breve{S}_n(w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i - \hat{G}^{(-i)}(X_i) - q]\hat{\ell}(W_i,w).$$

Define

$$S_{n4}(w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le g(X_i)] - q\}[\hat{\ell}(W_i,w) - \ell(W_i,w)],$$

$$S_{n5}(w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le G(X_i)] - I[Y_i \le g(X_i)]\}[\hat{\ell}(W_i,w) - \ell(W_i,w)],$$

and

$$S_{n6}(w) = n^{-1/2}\sum_{i=1}^{n}\{I[Y_i \le \hat{G}^{(-i)}(X_i)] - I[Y_i \le G(X_i)]\}[\hat{\ell}(W_i,w) - \ell(W_i,w)].$$

Then

$$\breve{S}_n(w) = \sum_{j=1}^{6} S_{nj}(w) .$$

Under assumptions 1-5 of Section 3.1 and assumptions 6-7 below, it follows from lemma A.3 of Horowitz and Lee (2009) that $S_{n4}(w) = o_p(1)$ uniformly over $w \in [0,1]$. Methods like those used to prove lemma 1 show that $S_{n6}(w) = o_p(1)$ uniformly over $w \in [0,1]$. Under H_0, $S_{n5}(w) = 0$, so the use of an estimated weight function does not affect Theorems 1 and 2. Theorem 3 is also unaffected because it is concerned with the behavior of $n^{-1}\tau_n$ as $n \to \infty$, and $n^{-1/2}S_{n5}(w) \to^p 0$ uniformly over $w \in [0,1]$ as $n \to \infty$. In addition, $S_{n5}(w) = o_p(1)$ uniformly over $w \in [0,1]$ under the sequence of local alternatives (3.3)-(3.4). Therefore, Theorem 4 is unaffected by estimation of ℓ.

Now consider Theorem 5. For any function $\delta(w,\eta)$, define

$$S_{n5}^*(\delta,\eta) = E\{I[Y \le G(X) - I[Y \le g(X)]\}\delta(W,\eta) .$$

Let

$$D_n(\eta) = S_{n3}(\eta) + n^{1/2}S_{n2}^*(\eta) + n^{1/2}S_{n5}^*[\hat{\ell}(W,\eta) - \ell(W,\eta),\eta] + S_{n6}(\eta) ,$$

and

$$\tilde{S}_n(\eta) = S_n(\eta) - D_n(\eta) .$$

As before,

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

Arguments like those used to prove theorem 5 and lemma 1 combined with $S_{n6}(\eta) = o_p(1)$ uniformly over $\eta \in [0,1]$ show that $\|\tilde{S}_n\|^2 = O_p(1)$. Therefore, as in the proof of Theorem 5,

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

and

$$\text{(6.11)} \quad P(\tau_n > z_\alpha) \ge P\left(0.5\|D_n\|^2 > z_\alpha + M\right)$$

for any sufficiently large M. But

$$S_{n5}^*(\hat{\ell} - \ell,\eta) = O\left(\|\pi\| \cdot \|\hat{\ell} - \ell\|\right)$$

and, under assumption 7(ii) below, $\|\pi\| \cdot \|\hat{\ell} - \ell\| / \|T\pi\| = o_p(1)$. Therefore, $\|S_{n5}^*(\hat{\ell} - \ell,\cdot)\| = o_p\left(\|T\pi\|\right)$. Now use $a^2 \ge 0.5b^2 - (b-a)^2$ with $a = D_n$ and $b = n^{1/2}S_{n2}^* + n^{1/2}S_{n5}^*$ to obtain,

$$\|D_n\|^2 \geq 0.5n\left\|S_{n2}^* + S_{n5}^*\right\|^2 - \|S_{n3} + S_{n6}\|^2 .$$

But $n^{1/2}S_{n2}^*(\eta) = n^{1/2}(T\pi)(\eta) + o_p(1)$, $\|T\pi\| > 0$, $\|S_{n3}\| = O_p(1)$, and $\|S_{n6}\| = o_p(1)$. Therefore,

(6.12) $\quad \|D_n\|^2 \geq Cn\|T\pi\|^2 + O_p(1)$

for all sufficiently large C. The theorem follows by substituting (6.12) into (6.11). Q.E.D.

The following are the additional assumptions needed to accommodate an estimated weight function.

Assumption 6:

(i) $\sup_{(\zeta,\eta)\in[0,1]^{p+r}} |\ell(z_2,w_2;\zeta,\eta) - \ell(z_1,w_1;\zeta,\eta)| \leq C_\ell \|(z_2,w_2)-(z_2,w_2)\|$ for each $(z,w)\in[0,1]^{p+r}$, where $\|(z_2,w_2)-(z_2,w_2)\|$ is the Euclidean distance between (z_2,w_2) and (z_1,w_1).

(ii) $\ell(z,w;\cdot,\cdot)\in C_{C_\ell}^{\nu}([0,1]^{p+r})$ for each $(z,w)\in[0,1]^{p+r}$ and some $\nu > (p+r)/2$.

Assumption 7:

(i) $\sup_{(z,w,\zeta,\eta)\in[0,1]^{2(p+r)}} |\hat{\ell}(z,w;\zeta,\eta) - \ell(z,w;\zeta,\eta)| = o_p(1)$ as $n\to\infty$.

(ii) $\sup_{g\in\mathcal{F}_{nC}} \|\pi\|\cdot\left\|\hat{\ell}-\ell\right\| / \|T\pi\| = o_p(1)$ as $n\to\infty$.

(iii) With probability approaching 1 as $n\to\infty$, $\sup_{(z,w,\zeta,\eta)\in[0,1]^{2(p+r)}} |\hat{\ell}(z,w;\zeta,\eta)| \leq C_\ell$,

$$\sup_{(\zeta,\eta)\in[0,1]^{p+r}} |\hat{\ell}(z_2,w_2;\zeta,\eta) - \hat{\ell}(z_1,w_1;\zeta,\eta)| \leq C_\ell \|(z_2,w_2)-(z_2,w_2)\|,$$

and for each $(z,w)\in[0,1]^{p+r}$, $\ell(z,w;\cdot,\cdot)\in C_{C_\ell}^{\nu}([0,1]^{p+r})$ for some $\nu > (p+r)/2$.

TABLE 1: RESULTS OF MONTE CARLO EXPERIMENTS WITH $\rho_1 = 0.35$

		Empirical Probability of Rejecting H_0			
n	ρ_2	τ_{nD}	τ_{nD}^*	τ_{nI}	τ_{nB}
750	0	0.093	0.073	0.056	0.074
	0.1	0.108	0.136	0.104	0.098
	0.2	0.194	0.334	0.279	0.184
	0.3	0.384	0.640	0.542	0.350
1000	0	0.072	0.067	0.053	0.072
	0.1	0.123	0.168	0.124	0.114
	0.2	0.226	0.416	0.328	0.230
	0.3	0.524	0.782	0.690	0.420
2000	0	0.062	0.056	0.046	0.054
	0.1	0.162	0.252	0.205	0.147
	0.2	0.483	0.697	0.608	0.384
	0.3	0.860	0.968	0.926	0.731

TABLE 2: RESULTS OF MONTE CARLO EXPERIMENTS WITH $\rho_1 = 0.7$

		Empirical Probability of Rejecting H_0			
n	ρ_2	τ_{nD}	τ_{nD}^*	τ_{nI}	τ_{nB}
750	0	0.063	0.048	0.050	0.098
	0.1	0.217	0.240	0.212	0.164
	0.2	0.640	0.757	0.660	0.410
	0.3	0.952	0.982	0.958	0.721
1000	0	0.057	0.048	0.054	0.086
	0.1	0.262	0.312	0.269	0.177
	0.2	0.610	0.861	0.794	0.511
	0.3	0.991	1.000	0.993	0.854
2000	0	0.054	0.044	0.049	0.056
	0.1	0.488	0.582	0.516	0.288
	0.2	0.980	0.996	0.985	0.840
	0.3	1.000	1.000	1.000	0.996

REFERENCES

Amemiya, T. (1982). Two stage least absolute deviations estimators, *Econometrica*, 50, 689-711.

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.

Blundell, R. and J.L. Horowitz (2007). A nonparametric test of exogeneity. *Review of Economic Studies*, 74, 1034-1058.

Blundell, R., J.L. Horowitz, and M. Parey (2015). Nonparametric estimation of a non-separable demand function under the Slutsky inequality restriction. Working paper, Department of Economics, Northwestern University.

Blundell, R. and J.L. Powell (2007). Censored regression quantiles with endogenous regressors, *Journal of Econometrics*, 141, 65-83.

Breunig, C. (2015). Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184, 328-346.

Chen, L. and S. Portnoy (1996). Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models, *Communications in Statistics, Theory and Methods*, 25, 1005-1032.

Chen, X. and D. Pouzo (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, 152, 46-60.

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80, 277-321.

Chen, X. and M. Reiss (2007). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*. 27:497-521.

Chernozhukov, V. and C. Hansen (2004). The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis, *Review of Economics and Statistics*, 86, 735-751.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects, *Econometrica*, 73, 245-261.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models, *Journal of Econometrics*, 132, 491-525.

Chernozhukov, V., G.W. Imbens, and W.K. Newey (2007). Instrumental variable identification and estimation of nonseparable models via quantile conditions, *Journal of Econometrics*, 139, 4-14.

Chesher, A. (2003). Identification in nonseparable models, *Econometrica*, 71, 1405-1441.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, 73, 1525-1550.

Chesher, A. (2007). Instrumental values. *Journal of Econometrics*, 139, 15-34.

Engl, H.W., M. Hanke, and A. Neubauer (1996). *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publishers.

Gasser, T. and H.G. Müller (1979). Kernel Estimation of Regression Functions, in *Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics*, 757, 23-68. New York: Springer.

Gasser, T. and H.G. Müller, and V. Mammitzsch (1985). Kernels and Nonparametric Curve Estimation, *Journal of the Royal Statistical Society Series B*, 47, 238-252.

Groetsch, C. (1984). *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. London: Pitman.

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 (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*. 33:-2904-2929.

Horowitz, J.L. and S. Lee (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75, 1191-1208.

Horowitz, J.L. and S. Lee (2009). Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152, 141-152.

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.

Januszewski, S.I. (2002). The effect of air traffic delays on airline prices, working paper, Department of Economics, University of California at San Diego, La Jolla, CA.

Kong, E., O. Linton, and Y. Xia (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application. *Econometric Theory*, 26, 1529-1564.

Lee, S. (2007): Endogeneity in quantile regression models: a control function approach, *Journal of Econometrics*, 141, 1131-1158.

Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

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

Ma, L. and R. Koenker (2006). Quantile regression methods for recursive structural equation models, *Journal of Econometrics*, 134, 471-506.

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

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57, 1027-1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Powell, J.L. (1983). The asymptotic normality of two-stage least absolute deviations estimators, *Econometrica*, 50, 1569-1575.

Sakata, S. (2007). Instrumental variable estimation based on conditional median restriction, *Journal of Econometrics*, 141, 350-382.

Song K. (2010). Testing semiparametric conditional moment restrictions using conditional martingale transforms. *Journal of Econometrics*, 154, 74-84.

Stute, W. and L. Shu (2005). Nonparametric checks for single-index models. *Annals of Statistics*, 33, 1048-1083.

van der Vaart, A.W. and J.A. Wellner (2007). *Empirical Processes Indexed by Estimated Functions*, IMS Lecture Notes-Monograph Series, 55, 234-252.

Yu, K. and M.C. Jones (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93, 228-237.