

**TESTING A PARAMETRIC QUANTILE-REGRESSION MODEL
WITH AN ENDOGENOUS EXPLANATORY VARIABLE AGAINST A
NONPARAMETRIC ALTERNATIVE***

Joel L. Horowitz

Department of Economics
Northwestern University
Evanston, IL, 60208-2600
USA

and

Sokbae Lee

Department of Economics
University College London
London, WC1E 6BT
United Kingdom

October 2007

Abstract

This paper is concerned with inference about a function g that is identified by a conditional quantile restriction involving instrumental variables. The paper presents a test of the hypothesis that g belongs to a finite-dimensional parametric family against a nonparametric alternative. The test is not subject to the ill-posed inverse problem of nonparametric instrumental variables estimation. Under mild conditions, the test is consistent against any alternative model. In large samples, its power is arbitrarily close to 1 uniformly over a class of alternatives whose distance from the null hypothesis is $O(n^{-1/2})$, where n is the sample size. Monte Carlo simulations illustrate the finite-sample performance of the test.

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

JEL Codes: C12, C14

*The research of Joel L. Horowitz was supported in part by NSF Grant 00352675. The work of both authors was supported in part by ESRC grant RES-000-22-0704 and by the Leverhulme Trust through its funding of the research programme "Evidence, Inference and Inquiry." We also gratefully acknowledge the financial support from the Leverhulme Trust and the Economic and Social Research Council for the Centre for Microdata Methods and Practice.

TESTING A PARAMETRIC QUANTILE-REGRESSION MODEL WITH AN ENDOGENOUS EXPLANATORY VARIABLE AGAINST A NONPARAMETRIC ALTERNATIVE

1 Introduction

Let Y , X , and W be random variables, and let g be a function that is identified by the relation

$$(1.1) \quad Y = g(X) + U; \quad P(U \leq 0 | W = w) = q$$

for some q satisfying $0 < q < 1$ and almost every w in the support of W . Equation (1.1) is a quantile-regression model in which Y is the dependent variable, X is a possibly endogenous explanatory variable, W is an instrument for X , and $U \equiv Y - g(X)$ is an unobserved random variable. This paper presents a test of the null hypothesis that g in (1.1) belongs to a finite-dimensional parametric family against a nonparametric alternative hypothesis. Specifically, let Θ be a compact subset of \mathbb{R}^d for some finite integer $d > 0$. The null hypothesis, H_0 , is that

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

for some $\theta \in \Theta$ and almost every x in the support of X , where G is a known function. The alternative hypothesis, H_1 , is that there is no θ such that (1.2) holds for almost every x . Under mild conditions, the test presented here is consistent against any alternative model. In large samples, its power is arbitrarily close to 1 uniformly over a class of alternative models whose “distance” from H_0 is $O(n^{-1/2})$, where n is the sample size.

Quantile regression models are increasingly important in applied econometrics. There has been much recent interest in nonparametric instrumental-variables (IV) estimation of quantile-regression models such as (1.1) and of models in which identification is achieved through the conditional mean restriction $E(U|W = w) = 0$. Chesher (2003, 2005, 2007); Chernozhukov and Hansen (2005); Chernozhukov, Imbens, and Newey (2007); and Horowitz and Lee (2007) discuss nonparametric identification and estimation of several versions of quantile-regression models with endogenous explanatory variables. Newey and Powell (2003); Darolles, Florens, and Renault (2006); Hall and Horowitz (2005); and Blundell, Chen, and Kristensen (2007) discuss nonparametric estimation of g under the conditional

mean restriction $E(U|W = w) = 0$. Newey, Powell and Vella (1999) present a “control function” approach to estimating g .

Methods for testing (1.2) against a nonparametric alternative under the conditional mean restriction $E(U|W = w) = 0$ have been developed by Donald, Imbens, and Newey (2003); Tripathi and Kitamura (2003); and Horowitz (2006). In addition, the test of a conditional mean function developed by Bierens (1990) and Bierens and Ploberger (1997) can be modified to provide a test of (1.2) under the restriction $E(U|W = w) = 0$ (Horowitz 2006). Horowitz and Spokoiny (2001, 2002) provide extensive references to other tests for conditional mean and quantile functions. However, we are unaware of any existing method for testing (1.2) against a nonparametric alternative under the quantile restriction $P(U \leq 0|W = w) = q$. This paper presents such a test.

An ability to test the hypothesis (1.2) for model (1.1) is important because results obtained with a misspecified parametric model can be highly misleading, whereas nonparametric IV estimation of g can be very imprecise. Methods for parametric estimation of quantile-regression models with endogenous regressors are well known. Estimators of linear quantile regression models with endogenous right-hand side variables are described by Amemiya (1982), Powell (1983), Chen and Portnoy (1996), Honoré and Hu (2004), Chernozhukov and Hansen (2006), Ma and Koenker (2006), Sakata (2007), and Lee (2007), among others. Chernozhukov and Hansen (2004) and Januszewski (2002) used such models in economic applications. Nonlinear parametric models can be estimated by the generalized method of moments (GMM). Parametric estimators typically have a $n^{-1/2}$ rate of convergence in probability but are subject to misspecification. Nonparametric estimation essentially eliminates the possibility of misspecifying g but, owing to the ill-posed-inverse problem of nonparametric IV estimation, nonparametric IV estimators can have very slow rates of convergence. The rate of convergence of a nonparametric IV estimator of g is always slower than $O(n^{-1/2})$ and, depending on the details of the distribution of (Y, X, W) , may be slower than $O(n^{-\varepsilon})$ for any $\varepsilon > 0$ (Hall and Horowitz 2005; Horowitz and Lee 2007). Consequently, parametric IV estimation is more attractive than nonparametric estimation if there is justification for believing that the parametric model is not seriously misspecified. This paper provides a way to test the specification of a parametric model.

The test developed here is not affected by the ill-posed inverse problem and, consequently, is more “precise” than any nonparametric estimator of g . Specifically, the test can detect a large class of nonparametric alternative models whose distance from the null-

hypothesis model is $O(n^{-1/2})$. It is not unusual in nonparametric estimation for rates of testing to be faster than rates of estimation. Nonparametric estimation and testing of conditional mean and median functions and nonparametric IV estimation and testing under the conditional moment restriction $E[U|W = w] = 0$ are other settings in which the rate of testing is faster than the rate of estimation. See Guerre and Lavergne (2002), Horowitz and Spokoiny (2001, 2002), and Horowitz (2006).

The test developed here builds on the results of Horowitz (2006), who developed a test of (1.2) against a nonparametric alternative under the conditional mean restriction $E(U|W = w) = 0$. Although there are similarities between the test presented here and that of Horowitz (2006), mean and quantile regressions are sufficiently different to require separate treatments. Nonparametric quantile IV produces an estimation problem that is nonlinear and non-smooth, whereas IV estimation under a conditional mean restriction has neither of these complications. Consequently, the methods that are needed to establish the properties of a test of (1.2) under a conditional quantile restriction are different from those that work under a conditional mean restriction. Specifically, our test statistic is a discontinuous function of an estimated parameter. See equation (2.3) in Section 2.1. The discontinuity of the test statistic greatly complicates the derivation of the test statistic's asymptotic distribution. We use empirical process methods to deal with the discontinuity. These methods are not needed for the test statistic of Horowitz (2006), which is a continuous function of the estimated parameter and, therefore, much easier to analyze.

Section 2 describes the test statistic and its properties. Section 3 presents the results of a Monte Carlo investigation of the finite-sample performance of the test. Section 4 presents concluding comments. The proofs of theorems are in the mathematical appendix, which is Section 5.

2 The Test Statistic and Its Asymptotic Properties

Assume for now that Y , X , and W are continuously distributed, scalar random variables with joint probability density function f_{YXW} . The extension to the case in which X and W are vectors and some components of X may be exogenous is straightforward and is outlined in Section 2.10. Assume, also, that the supports of X and W are contained in $[0, 1]$. This assumption can always be satisfied by carrying out monotone transformations of X and W . The data, $\{(Y_i, X_i, W_i) : i = 1, \dots, n\}$, are a simple random sample of (Y, X, W) .

2.1 The Test Statistic

Equation (1.1) implies that

$$P[Y - g(X) \leq 0 | W = w] - q = 0$$

for almost every $w \in [0, 1]$. Under H_0 , $g(x) = G(x, \theta)$ for almost every $x \in [0, 1]$, some unknown $\theta \in \Theta$, and a known function G . Therefore, H_0 is equivalent to

$$(2.1) \quad P[Y - G(X, \theta) \leq 0 | W = w] - q = 0.$$

Let f_W denote the probability density function of W . Define

$$F_{YXW}(y, x, w) = \int_{-\infty}^y f_{YXW}(v, x, w) dv$$

and

$$\tilde{S}(w) = \int_0^1 \{F_{YXW}[G(x, \theta), x, w] - qf_W(w)\} dx.$$

Then (2.1) is equivalent to

$$(2.2) \quad \tilde{S}(w) = 0$$

for almost every $w \in [0, 1]$. H_1 is equivalent to the statement that there is no $\theta \in \Theta$ such that (2.2) holds for almost every $w \in [0, 1]$. A test of H_0 can be based on a sample analog of

$$\int_0^1 \tilde{S}(w)^2 dw,$$

but the resulting rate of testing is slower than $n^{-1/2}$. A rate of $n^{-1/2}$ can be achieved by smoothing \tilde{S} . To this end, let $l(z, w)$ denote the kernel of a nonsingular integral operator, L , on $L_2[0, 1]$. That is, the operator L is defined by

$$L\varphi(z) = \int_0^1 l(z, w)\varphi(w)dw.$$

Because L is nonsingular, H_0 is equivalent to

$$\begin{aligned} S(z) &\equiv (L\tilde{S})(z) \\ &= \int_0^1 \int_0^1 \{F_{YXW}[G(x, \theta), x, w] - qf_W(w)\} l(z, w) dx dw \\ &= 0 \end{aligned}$$

for almost every $z \in [0, 1]$ and for some $\theta \in \Theta$. H_1 is equivalent to the statement that there is no $\theta \in \Theta$ such that $S(z) = 0$ for almost every $z \in [0, 1]$. Define

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

Then H_0 is true if and only if $\tau = 0$. The test statistic developed here is a sample analog of τ .

To form the analog, let $\hat{\theta}$ be an estimator of θ that is consistent under H_0 . For reasons that are explained in Section 2.8, it is convenient to permit l to depend on the distribution of (Y, X, W) so that l must be estimated from the data. Let \hat{l} be a consistent estimator of l , which can be l itself if l does not depend on the distribution of (Y, X, W) . The sample analog of $S(z)$ is

$$(2.3) \quad S_n(z) = n^{-1/2} \sum_{i=1}^n \left[I \left\{ Y_i \leq G(X_i, \hat{\theta}) \right\} - q \right] \hat{l}(z, W_i),$$

where $I(\cdot)$ is the indicator function. The test statistic is

$$(2.4) \quad \tau_n = \int_0^1 S_n^2(z) dz.$$

H_0 is rejected if τ_n is large. The asymptotic distribution of τ_n under H_0 is given in Theorem 1 in Section 2.3. Section 2.4 presents a method for computing the critical value.

2.2 Regularity Conditions

This section states the assumptions that are used to obtain the asymptotic properties of τ_n under the null and alternative hypotheses.

Assumption 1. (i) The support of (X, W) is $[0, 1]^2$. (ii) (Y, X, W) has a probability density function f_{YXW} with respect to Lebesgue measure. (iii) f_{YXW} is bounded and is differentiable with respect to its first argument with a uniformly bounded derivative. (iv) $\{(Y_i, X_i, W_i) : i = 1, \dots, n\}$ is a simple random sample of (Y, X, W) .

Assumption 2. There is a constant $C_G < \infty$ such that $\sup_{x \in [0, 1]} |g(x)| \leq C_G$, $\sup_{\theta \in \Theta} \sup_{x \in [0, 1]} |G(x, \theta)| \leq C_G$, and the first and second derivatives of $G(x, \theta)$ with respect to θ are bounded by C_G uniformly over $x \in [0, 1]$ and $\theta \in \Theta$.

Assumption 3. (i) As $n \rightarrow \infty$, $\hat{\theta} \rightarrow_p \theta_0$ for some $\theta_0 \in \Theta$, a compact subset of \mathbb{R}^d . (ii) If H_0 is true, then $g(x) \equiv G(x, \theta_0)$, $\theta_0 \in \text{int}(\Theta)$, and

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

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

Assumption 4. (i) $\sup_{(z,w) \in [0,1]^2} |\hat{l}(z, w) - l(z, w)| = o_p(1)$. (ii) The operator L is nonsingular. (iii) There is a constant $C_l < \infty$ such that

$$\begin{aligned} \sup_{(z,w) \in [0,1]^2} |l(z, w)| &\leq C_l, \\ \sup_{w \in [0,1]} |l(z_1, w) - l(z_2, w)| &\leq C_l |z_1 - z_2|, \\ \sup_{z \in [0,1]} |l(z, w_1) - l(z, w_2)| &\leq C_l |w_1 - w_2|, \end{aligned}$$

and with probability approaching 1 as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{(z,w) \in [0,1]^2} |\hat{l}(z, w)| &\leq C_l, \\ \sup_{w \in [0,1]} |\hat{l}(z_1, w) - \hat{l}(z_2, w)| &\leq C_l |z_1 - z_2|, \\ \sup_{z \in [0,1]} |\hat{l}(z, w_1) - \hat{l}(z, w_2)| &\leq C_l |w_1 - w_2|. \end{aligned}$$

Assumption 1 specifies properties of the distribution of the data. Assumption 2 places mild boundedness and smoothness restrictions on g and G . Assumption 3 is satisfied, for example, by the GMM estimator of θ_0 that is defined in Section 2.4. Assumption 4 can be satisfied by making suitable choices of l and \hat{l} . The choices of l and \hat{l} are discussed further in Sections 2.8-2.9.

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

To obtain the asymptotic distribution of τ_n under H_0 , define $G_\theta(x, \theta) = \partial G(x, \theta) / \partial \theta$,

$$\begin{aligned} \Gamma(z) &= \int_0^1 \int_0^1 f_{YXW}[G(x, \theta_0), x, w] l(z, w) G_\theta(x, \theta_0) dx dw, \\ B_n(z) &= n^{-1/2} \sum_{i=1}^n [\{I[Y_i \leq g(X_i)] - q\} l(z, W_i) + \Gamma(z)' \gamma(Y_i, X_i, W_i, \theta_0)], \end{aligned}$$

and

$$V(z_1, z_2) = E [B_n(z_1)B_n(z_2)].$$

Define the operator Ω on $L_2[0, 1]$ by

$$(2.5) \quad (\Omega\psi)(z) = \int_0^1 V(z, \zeta)\psi(\zeta)d\zeta.$$

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

Theorem 1. *Let Assumptions 1-4 hold. If H_0 is true, then*

$$\tau_n \rightarrow_d \sum_{j=1}^{\infty} \omega_j \chi_{1j}^2.$$

2.4 Obtaining the Critical Value

The statistic τ_n is not asymptotically pivotal, so its asymptotic distribution cannot be tabulated. This section presents a method for obtaining an approximate asymptotic critical value for the τ_n test using a pseudo-true model, as in Horowitz (2006). Let $Q(A|B)$ denote the q -th quantile of a random variable A conditional on B and define a model

$$(2.6) \quad \tilde{Y} = G(X, \theta_0) + \tilde{U},$$

where $\tilde{Y} = Y - Q[Y - G(X, \theta_0)|W]$, $\tilde{U} = \tilde{Y} - G(X, \theta_0)$, and θ_0 is the probability limit of $\hat{\theta}$. Note that this model coincides with (1.1) if H_0 is true, since $Q[Y - G(X, \theta_0)|W] = 0$ under H_0 . Furthermore, even if H_0 is false, $Q[\tilde{Y} - G(X, \theta_0)|W] = 0$, thereby implying that H_0 holds for the pseudo-true model (2.6) whether or not H_0 holds for (1.1). The approximate critical value for the τ_n test is obtained from the asymptotic distribution of τ_n under sampling from model (2.6).

Let $\{\tilde{\omega}_j : j = 1, 2, \dots\}$ denote the eigenvalues of the version of Ω (say, $\tilde{\Omega}$) that is obtained by replacing model (1.1) with model (2.6). As before, order the eigenvalues such that $\tilde{\omega}_1 \geq \tilde{\omega}_2 \geq \dots \geq 0$. Then under sampling from (2.6), τ_n is asymptotically distributed as

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

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

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

uniformly over t . Define

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

Let $z_{\varepsilon\alpha}$ denote the $1 - \alpha$ quantile of the distribution of $\tilde{\tau}_\varepsilon$. Then

$$0 < P(\tilde{\tau} > z_{\varepsilon\alpha}) - \alpha < \varepsilon.$$

Thus, using $z_{\varepsilon\alpha}$ to approximate the asymptotic α -level critical value of τ_n creates an arbitrarily small error in the probability that a correct H_0 is rejected. Similarly, use of the approximation creates an arbitrarily small change in the power of the τ_n test when H_0 is false. However, the eigenvalues $\tilde{\omega}_j$ are unknown. Accordingly, the approximate α -level critical value for the τ_n test is the consistent estimator of the $1 - \alpha$ quantile of the distribution of $\tilde{\tau}_\varepsilon$ that is obtained by replacing the unknown eigenvalues with consistent estimates $\hat{\omega}_j$, ($j = 1, \dots, K_\varepsilon$). That is, the approximate value of τ_n is the $1 - \alpha$ quantile of the distribution of

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

We now describe how to obtain the estimated eigenvalues $\tilde{\omega}_j$. To do so, let \mathcal{W} be a d -vector of instruments for X that is derived from W (for example, powers of W). Let \mathcal{W}_i be the value of \mathcal{W} that corresponds to W_i . Assume that $\hat{\theta}$ satisfies

$$(2.7) \quad n^{-1/2} \sum_{i=1}^n \mathcal{W}_i \left[I \left\{ Y_i \leq G(X_i, \hat{\theta}) \right\} - q \right] = o_p(1)$$

as $n \rightarrow \infty$. This relationship is satisfied, for example, by a GMM estimator with instruments $\{\mathcal{W}_i : i = 1, \dots, n\}$ and by the IV quantile-regression estimator of Chernozhukov and Hansen (2006). Let $f_{U|XW}$ denote the density function of U in (1.1) conditional on (X, W) . Define the $d \times d$ matrix

$$\Phi = E \left[f_{U|X,W}(0|X, W) \mathcal{W} G_\theta(X, \theta_0)' \right],$$

and define

$$(2.8) \quad \tilde{\gamma}(Y_i, X_i, W_i, \theta_0) = \Phi^{-1} \mathcal{W}_i [I \{Y_i \leq G(X_i, \theta_0)\} - q].$$

It can be shown that $\gamma(Y_i, X_i, W_i, \theta_0)$ in Assumption 3 is $\gamma(Y_i, X_i, W_i, \theta_0) = -\tilde{\gamma}(Y_i, X_i, W_i, \theta_0)$. Then

$$V(z_1, z_2) = q(1 - q)E \left[\{l(z_1, W) - \Gamma(z_1)' \Phi^{-1} \mathcal{W}\} \{l(z_2, W) - \Gamma(z_2)' \Phi^{-1} \mathcal{W}\} \right].$$

To construct a consistent estimator of V , let K denote a kernel function with a bandwidth δ_n . Let $\hat{Y}_i = Y_i - \hat{Q}[Y - G(X, \hat{\theta})|W_i]$, where $\hat{Q}[\cdot|W]$ denotes a nonparametric estimator of $Q[\cdot|W]$, e.g. a local linear quantile regression estimator. Define

$$(2.9) \quad \hat{\Gamma}(z) = (n\delta_n)^{-1} \sum_{i=1}^n K \left(\frac{\hat{Y}_i - G(X_i, \hat{\theta})}{\delta_n} \right) \hat{l}(z, W_i) G_{\theta}(X_i, \hat{\theta})$$

and

$$(2.10) \quad \hat{\Phi} = (n\delta_n)^{-1} \sum_{i=1}^n K \left(\frac{\hat{Y}_i - G(X_i, \hat{\theta})}{\delta_n} \right) \mathcal{W}_i G_{\theta}(X_i, \hat{\theta})'.$$

Under regularity conditions, it is straightforward to show that $\hat{\Gamma}(z)$ and $\hat{\Phi}$ are consistent estimators of $\Gamma(z)$ and Φ . Then $V(z_1, z_2)$ can be estimated consistently by

$$\hat{V}(z_1, z_2) = q(1 - q)n^{-1} \sum_{i=1}^n \left[\left\{ \hat{l}(z_1, W_i) - \hat{\Gamma}(z_1)' \hat{\Phi}^{-1} \mathcal{W}_i \right\} \left\{ \hat{l}(z_2, W_i) - \hat{\Gamma}(z_2)' \hat{\Phi}^{-1} \mathcal{W}_i \right\} \right].$$

Let $\hat{\Omega}$ be the integral operator whose kernel is $\hat{V}(z_1, z_2)$ and let $\hat{\omega}_j$ be the eigenvalues of $\hat{\Omega}$. Also, let $\hat{z}_{\varepsilon\alpha}$ denote the $1 - \alpha$ quantile of the distribution of $\hat{\tau}_n$. Theorem 2 gives conditions under which the $\hat{\omega}_j$'s are consistent for the $\tilde{\omega}_j$'s and $\hat{z}_{\varepsilon\alpha}$ is consistent for $z_{\varepsilon\alpha}$.

Theorem 2. *Let Assumptions 1-4 hold. Assume that K has support $[-1, 1]$, is bounded, Lipschitz continuous, continuously differentiable, and symmetrical about 0, and $\int_{-1}^1 K(u)du = 1$. Assume that $Q[Y - G(X, \theta)|W = w]$ is continuously differentiable with respect to w . In addition, assume that $\delta_n \rightarrow 0$, $\log n/(n\delta_n) \rightarrow 0$, and $\delta_n^{-1}n^{-1/3}(\log n)^{1/2} \rightarrow 0$. Then as $n \rightarrow \infty$, (i) $\sup_{1 \leq j \leq K_{\varepsilon}} |\hat{\omega}_j - \tilde{\omega}_j| = o_p(1)$ and (ii) $\hat{z}_{\varepsilon\alpha} \rightarrow_p z_{\varepsilon\alpha}$.*

We now describe how to obtain an accurate numerical approximation to the $\hat{\omega}_j$'s. Let $\hat{L}(z)$ denote the $n \times 1$ vector whose i th component is $\hat{l}(z, W_i)$, let \hat{F} denote the $n \times d$ matrix whose i th row is

$$\frac{1}{\delta_n} K \left(\frac{\hat{Y}_i - G(X_i, \hat{\theta})}{\delta_n} \right) G_{\theta}(X_i, \hat{\theta})',$$

and let $\tilde{\mathcal{W}}$ denote $n \times d$ matrix whose i th row is \mathcal{W}'_i . In addition, let $\hat{M} = I_n - n^{-1}\hat{F}\hat{\Phi}^{-1}\tilde{\mathcal{W}}'$, where I_n is the $n \times n$ identity matrix. Then

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

To obtain a finite-dimensional approximation to the $\hat{\omega}_j$'s, let $\{\psi_j : j = 1, 2, \dots\}$ be an orthonormal basis for $L_2[0, 1]$. Then

$$\hat{l}(z, w) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \hat{d}_{jk} \psi_j(z) \psi_k(w),$$

where

$$(2.11) \quad \hat{d}_{jk} = \int_0^1 \int_0^1 \hat{l}(z, w) \psi_j(z) \psi_k(w) dz dw.$$

Approximate $\hat{l}(z, w)$ by the finite sum

$$\Pi(z, w) = \sum_{j=1}^J \sum_{k=1}^J \hat{d}_{jk} \psi_j(z) \psi_k(w)$$

for some finite integer $J \geq K_\varepsilon$, which can be chosen large enough to make Π approximate \hat{l} with any desired accuracy. Let $\psi(z)$ denote the $J \times 1$ vector whose j th component is $\psi_j(z)$ and Ψ denote the $J \times n$ matrix whose (j, k) th component is $\psi_j(W_k)$. Let \hat{D} be the $J \times J$ matrix $\{\hat{d}_{jk}\}$. Then $\hat{V}(z_1, z_2)$ is approximated by

$$\hat{V}(z_1, z_2) = q(1 - q)n^{-1}\psi(z_1)'\hat{D}\Psi\hat{M}\hat{M}'\Psi'\hat{D}'\psi(z_2).$$

The eigenvalues of $\hat{\Omega}$ are approximated by those of the $J \times J$ matrix $q(1 - q)n^{-1}\hat{D}\Psi\hat{M}\hat{M}'\Psi'\hat{D}'$.

2.5 Consistency of the Test against a Fixed Alternative Model

In this section, it is assumed that H_0 is false; that is, there is no $\theta \in \Theta$ such that $g(x) = G(x, \theta)$ for almost every $x \in [0, 1]$. Let θ_0 denote the probability limit of $\hat{\theta}_n$. Define

$$\mathcal{T}(z) = \int_0^1 \int_0^1 \{F_{YXW}[G(x, \theta_0), x, w] - F_{YXW}[g(x), x, w]\} l(z, w) dx dw.$$

Let \tilde{z}_α denote the $1 - \alpha$ quantile of the distribution of τ_n under sampling from the pseudo-true model (2.6). Let $\hat{z}_{\varepsilon\alpha}$ denote the $1 - \alpha$ quantile of the distribution of $\hat{\tau}_n$. The following theorem establishes consistency of the τ_n test against a fixed alternative hypothesis.

Theorem 3. *Let Assumptions 1-4 hold. If H_0 is false and $\int_0^1 \mathcal{T}^2(z)dz > 0$, then*

$$\lim_{n \rightarrow \infty} Pr(\tau_n > \tilde{z}_\alpha) = 1$$

and

$$\lim_{n \rightarrow \infty} Pr(\tau_n > \hat{z}_{\varepsilon\alpha}) = 1.$$

If g is identified, then $\int_0^1 \mathcal{T}^2(z)dz = 0$ only if $G(x, \theta_0) = g(x)$ for almost every $x \in [0, 1]$ (Horowitz and Lee, 2007). Therefore, the τ_n test is consistent against any identified alternative model that differs from $G(x, \theta_0)$ on a set of x values whose Lebesgue measure exceeds 0.

2.6 Asymptotic Distribution under Local Alternatives

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

$$(2.12) \quad Y = G(X, \theta_0) + n^{-1/2}\Delta(X) + U, \quad P(U \leq 0 | W = w) = q$$

for almost every $w \in [0, 1]$, where Δ is a bounded function on $[0, 1]$ and $\theta_0 \in \text{int}(\Theta)$.

To obtain the asymptotic distribution, assume that $\hat{\theta}$ satisfies (2.7). Let $\{(\omega_j, \phi_j) : j = 1, 2, \dots\}$ denote the eigenvalues and orthonormal eigenvectors of the version of the operator Ω in (2.5) that is obtained by setting

$$\gamma(Y_i, X_i, W_i, \theta_0) = -\tilde{\gamma}(Y_i, X_i, W_i, \theta_0),$$

where $\tilde{\gamma}$ is defined in (2.8). Order ω_j 's so that $\omega_1 \geq \omega_2 \geq \dots \geq 0$. In addition, define

$$\mu(z) = \Gamma(z)' \Phi^{-1} E [f_{U|X,W}(0|X, W)\Delta(X)W] - E [f_{U|X,W}(0|X, W)\Delta(X)l(z, W)]$$

and

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

Let $\{\chi_{1j}^2(\mu_j^2/\omega_j) : j = 1, 2, \dots\}$ denote independent random variables that are distributed as noncentral chi-square with one degree of freedom and noncentrality parameters $\{\mu_j^2/\omega_j\}$.

The following theorem states the result.

Theorem 4. *Let Assumptions 1-4 hold. Assume that $\hat{\theta}$ satisfies (2.7). Under the sequence of local alternative hypotheses (2.12),*

$$\tau_n \rightarrow_d \sum_{j=1}^{\infty} \omega_j \chi_{1j}^2(\mu_j^2/\omega_j).$$

Let z_α denote the $1 - \alpha$ quantile of the distribution of $\sum_{j=1}^{\infty} \omega_j \chi_{1j}^2(\mu_j^2/\omega_j)$. Let $\hat{z}_{\varepsilon\alpha}$ denote the estimated approximate α -level critical value defined in Section 2.3. Then it follows from Theorems 2 and 4 that for any $\varepsilon > 0$

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

It also follows from Theorem 4 that the τ_n test has power against local alternatives whose distance from the null-hypothesis model is $O(n^{-1/2})$. If $\mu(z) = 0$ for all $z \in [0, 1]$, then there is a non-stochastic sequence $\{\theta_n\}$ such that

$$G(x, \theta_n) = G(x, \theta) + n^{-1/2} \Delta(x) + o(n^{-1/2}).$$

Therefore, the distance between the null and alternative hypotheses is $o(n^{-1/2})$.

2.7 Uniform Consistency

This section shows that for any $\varepsilon > 0$, the τ_n test rejects H_0 with probability exceeding $1 - \varepsilon$ uniformly over a class of alternative models whose distance from the null hypothesis is $O(n^{-1/2})$. Uniform consistency is important because it provides some assurance that there are not alternatives against which a test has low power even with large samples. If a test is not uniformly consistent over a specified set, then that set contains alternatives against which the test has low power.

Let θ_g denote the probability limit of $\hat{\theta}$ under the hypothesis (not necessarily true) that $g(x) = G(x, \theta)$ for some $\theta \in \Theta$ and a given function G . Define $Q_g(x) = G(x, \theta_g) - g(x)$ and

$$\mathcal{T}_g(z) = \int_0^1 \int_0^1 \{F_{YXW}[G(x, \theta_g), x, w] - F_{YXW}[g(x), x, w]\} l(z, w) dx dw.$$

Define the set of functions $\mathcal{C}_{C_g}^\alpha(\mathcal{X})$ as follows. Let $\underline{\alpha}$ denote the greatest integer strictly smaller than α . For any vector $k = (k_1, \dots, k_d)$ of d integers, let D^k denote the differential operator

$$D^k = \frac{\partial^k}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} \quad \text{with } k. = \sum_{i=1}^d k_i.$$

In addition, let

$$\|g\|_\alpha = \max_{k \leq \alpha} \sup_x |D^k g(x)| + \max_{k=\alpha} \sup_{x,y} \frac{|D^k g(x) - D^k g(y)|}{\|x - y\|^{\alpha - k}},$$

where the suprema are taken over all x, y in the interior of \mathcal{X} with $x \neq y$. Then $\mathcal{C}_{C_g}^\alpha(\mathcal{X})$ is the set of all continuous functions $g : \mathcal{X} \subset \mathbf{R}^d \mapsto \mathbf{R}$ with $\|g\|_\alpha \leq C_g$. This class of smooth functions is used in Van der Vaart and Wellner (1996, p.154), Chen, Linton and Van Keilegom (2003), and Ichimura and Lee (2006), among others.

Let $\tilde{\Theta}$ be a compact subset of $\text{int}(\Theta)$. Let $\|\cdot\|$ denote the $L_2[0, 1]$ norm. For each $n = 1, 2, \dots$ and $C > 0$, define \mathcal{F}_{nc} as a set of functions g such that (i) $g \in \mathcal{C}_{C_g}^\alpha([0, 1])$ for some $\alpha > 1$ and some constant $C_g < \infty$, (ii) $\theta_g \in \tilde{\Theta}$, (iii) $n^{1/2}(\hat{\theta} - \theta_g) = O_p(1)$ uniformly over $g \in \mathcal{F}_{nc}$, (iv) $\|\mathcal{T}_g\| \geq n^{-1/2}C$, and (v) $\sup_{g \in \mathcal{F}_{nc}} \|Q_g\| \left\| \hat{l} - l \right\| / \|\mathcal{T}_g\| = o_p(1)$ as $n \rightarrow \infty$. Condition (ii) ensures the existence of the critical value defined in Section 2.3. The condition is not restrictive in applications because Θ and $\tilde{\Theta}$ can usually be made large enough to include any reasonable θ_g . Condition (iv) implies that \mathcal{F}_{nc} includes alternative models for which $\|Q_g\| = O(n^{-1/2})$. To understand condition (v), define the operator T_g by

$$(T_g \psi)(z) = \int_0^1 \int_0^1 f_{YXW}[g(x), x, w] l(z, w) \psi(x) dx dw.$$

Then $\mathcal{T}_g = T_g Q_g + O(\|Q_g\|^2)$. Condition (v) rules out deviations $Q_g(x)$ from the null hypothesis that depend on x only through sequences of eigenvectors of T_g whose eigenvalues converge to 0 too rapidly. The practical significance of this condition is that the τ_n test has relatively low power against alternatives that differ from H_0 only through eigenvectors of T_g with very small eigenvalues.

The following theorem states the result of this section.

Theorem 5. *Let Assumptions 1, 2, and 4 hold. Assume that $\hat{\theta}$ satisfies (2.7). Then for any $\delta > 0$ and α such that $0 < \alpha < 1$, and sufficiently large but finite constant C ,*

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{F}_{nc}} Pr(\tau_n > z_\alpha) \geq 1 - \delta$$

and

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

2.8 Weight Functions

This section considers the choice of the weight function $l(z, w)$. We show that setting $l(z, w) = f_{YXW}[g(z), z, w]$ has certain power advantages over a weight function that does not depend on the distribution of (Y, X, W) . Section 2.9 presents a method for estimating $f_{YXW}[g(z), z, w]$.

Let τ_{ng} denote the test statistic with $l(z, w) = f_{YXW}[g(z), z, w]$ and τ_{nl} denote the statistic with a weight function that is independent of the distribution of (Y, X, W) . We show that the power of the τ_{nl} test can be low relative to that of the τ_{ng} test. Specifically, there are combinations of density functions f_{YXW} and local alternative models such that an α -level τ_{nl} test based on a fixed weight function has asymptotic local power that is arbitrarily close to α , whereas the asymptotic local power of the α -level τ_{ng} test is bounded away from and above α . The opposite situation cannot occur under the assumptions of this paper. That is, it is not possible for the asymptotic local power of the α -level τ_{ng} test to approach α while the power of the α -level τ_{nl} test remains bounded away from α .

The conclusion that the power of τ_{nl} can be low relative to that of τ_{ng} is reached by constructing an example in which the α -level test has asymptotic power that is bounded away from α but the τ_{nl} test has asymptotic power that is arbitrarily close to α . To minimize the complexity of the example, assume that θ_0 is known a priori and does not have to be estimated. Define

$$\begin{aligned}\bar{B}_{ng}(z) &= n^{-1/2} \sum_{i=1}^n [I\{Y_i \leq g(X_i)\} - q] f_{YXW}[g(z), z, W_i], \\ \bar{B}_{nl}(z) &= n^{-1/2} \sum_{i=1}^n [I\{Y_i \leq g(X_i)\} - q] l(z, W_i),\end{aligned}$$

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

$$\begin{aligned}(\bar{\Omega}_g\psi)(z) &= \int_0^1 \bar{R}_g(z, \xi)\psi(\xi)d\xi \\ (\bar{\Omega}_l\psi)(z) &= \int_0^1 \bar{R}_l(z, \xi)\psi(\xi)d\xi.\end{aligned}$$

Let $\{(\bar{\omega}_{jg}, \bar{\psi}_{jg}) : j = 1, 2, \dots\}$ and $\{(\bar{\omega}_{jl}, \bar{\psi}_{jl}) : j = 1, 2, \dots\}$ denote the eigenvalues and eigenvectors of $\bar{\Omega}_g$ and $\bar{\Omega}_l$, respectively, sorted in decreasing order. For Δ defined as in

(2.12), define

$$\bar{\mu}_g(z) = - \int_0^1 \int_0^1 f_{YXW}(g(x), x, w) \Delta(x) f_{YXW}(g(z), z, w) dx dw,$$

$$\bar{\mu}_l(z) = - \int_0^1 \int_0^1 f_{YXW}(g(x), x, w) \Delta(x) l(z, w) dx dw,$$

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

and

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

Then arguments identical to those used to prove Theorem 4 yields that under the sequence of local alternative hypotheses (2.12) with a known θ_0 ,

$$\tau_{ng} \rightarrow_d \sum_{j=1}^{\infty} \bar{\omega}_{jg} \chi_{1j}^2(\bar{\mu}_{jg}^2 / \bar{\omega}_{jg}),$$

and

$$\tau_{nl} \rightarrow_d \sum_{j=1}^{\infty} \bar{\omega}_{jl} \chi_{1j}^2(\bar{\mu}_{jl}^2 / \bar{\omega}_{jl})$$

as $n \rightarrow \infty$.

Therefore, to establish the first conclusion of this section, it suffices to show that for any fixed function l , f_{YXW} and Δ can be chosen so that $\|\bar{\mu}_g\|^2 / \sum_{j=1}^{\infty} \bar{\omega}_{jg}$ is bounded away from 0 and $\|\bar{\mu}_l\|^2 / \sum_{j=1}^{\infty} \bar{\omega}_{jl}$ 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 $m > 1$ be a finite integer. Define

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

Let

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

Then

$$\bar{R}_g(z_1, z_2) = q(1 - q) E_W [f_{UXW}(0, z_1, W) f_{UXW}(0, z_2, W)],$$

$\bar{\omega}_{jg} = \lambda_j$ and $\sum_{j=1}^{\infty} \bar{\omega}_{jg}$ is non-zero and finite. Set $\Delta(x) = D\phi_m(x)$ for some finite $D > 0$. Then $\|\bar{\mu}_g\|^2 = D^2 \lambda_m^2 = D^2$. It suffices to show that m can be chosen so that $\|\bar{\mu}_l\|$ is

arbitrarily close to 0. To do this, note that $l(z, w)$ has the Fourier representation

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

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

We next show that the opposite situation cannot occur. That is, we show below that there exists a universal constant C such that

$$(2.13) \quad \|\bar{\mu}_l\| \leq C \|\bar{\mu}_g\|$$

for any l , provided that

$$\int_0^1 \int_0^1 [l(x, w)]^2 dx dw < C_l$$

and $\|\Delta\|^2 < C_{\Delta}$ for some constants $C_l < \infty$ and $C_{\Delta} < \infty$. To show (2.13), use the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} \|\bar{\mu}_l\|^2 &\leq \left[\int_0^1 \int_0^1 [l(x, w)]^2 dx dw \right] \int_0^1 \left[\int_0^1 f_{YXW}(g(x), x, w) \Delta(x) dx \right]^2 dw \\ &\leq C_l \int_0^1 \left[\int_0^1 f_{YXW}(g(x), x, w) \Delta(x) dx \right]^2 dw \\ &= C_l \int_0^1 \int_0^1 \left[\int_0^1 f_{YXW}(g(x), x, w) f_{YXW}(g(z), z, w) dw \right] \Delta(x) dx \Delta(z) dz \\ &= C_l \int_0^1 |\bar{\mu}_g(z)| \Delta(z) dz \\ &\leq C_l \|\bar{\mu}_g\|^2 \|\Delta\|^2 \\ &\leq C_l C_{\Delta} \|\bar{\mu}_g\|^2, \end{aligned}$$

which proves (2.13). Therefore, $\|\bar{\mu}_g\|^2$ can approach 0 only if $\|\bar{\mu}_l\|^2$ also approaches 0.

2.9 Estimating the Weight Function

We now explain how to estimate the weight function $l(z, w) = f_{YXW}[g(z), z, w]$. Let \hat{g} denote Horowitz's and Lee's (2007) nonparametric estimator of g in (1.1). Let $\hat{f}_{YXW}(y, x, w)$ denote a kernel nonparametric estimator of $f_{YXW}(y, x, w)$. Then $\hat{f}_{YXW}[\hat{g}(z), z, w]$ is an

obvious estimator of $f_{YXW}[g(z), z, w]$, but this estimator does not satisfy assumption 4 of Section 2.2. We overcome this problem by using a kernel to smooth \hat{g} .

Let K be a symmetrical, continuously differentiable probability density function on $[-1, 1]$. Let $\{h_n : n = 1, 2, \dots\}$ be a sequence of strictly positive constants such that $h_n^{-2} \|\hat{g} - g\| = o_p(1)$. Define smoothed versions of g and \hat{g} by

$$\tilde{g}(z) = h_n^{-1} \int_0^1 K\left(\frac{x-z}{h_n}\right) g(x) dx$$

and

$$\tilde{\hat{g}}(z) = h_n^{-1} \int_0^1 K\left(\frac{x-z}{h_n}\right) \hat{g}(x) dx.$$

Also, define

$$\tau_{nh} = \int_{1-h_n}^{h_n} S_{nh}^2(z) dz,$$

where S_{nh} is the version of S_n that is obtained by setting $\hat{l}(z, w) = \hat{f}_{YXW}(\tilde{\hat{g}}(z), z, w)$ on the right-hand side of (2.3). Then τ_{nh} can be used in place of τ_n to test H_0 . As with τ_n , H_0 is rejected if τ_{nh} is large.

We show in Section 5 that

- (i) The conclusions of Theorems 1 and 3 hold for τ_{nh} .
- (ii) The test based on τ_{nh} rejects a false H_0 with probability greater than or equal to $1 - \delta$ for any $\delta > 0$ uniformly over a class of alternative models whose “distance” from H_0 on $[h_n, 1 - h_n]$ is $O(n^{-1/2})$.
- (iii) Inequality (2.13) holds if τ_{nh} is used in place of τ_n and n is sufficiently large.

Since $h_n \rightarrow 0$ as $n \rightarrow \infty$, these results imply that τ_{nh} can be used in place of τ_n in large samples.

2.10 Multivariate Extension

We now extend the τ_n test to the multivariate model

$$(2.14) \quad Y = g(X, Z) + U; \quad P(U \leq 0 | Z = z, W = w) = q$$

for for some q satisfying $0 < q < 1$ and almost every (z, w) , where Y and U are scalar random variables, X and W are random variables whose supports are contained in $[0, 1]^{p_x}$ and

$[0, 1]^{p_w}$, respectively, ($p_w \geq p_x \geq 1$), and Z is a random variable whose support is contained in $[0, 1]^r$ ($r \geq 0$). If $r = 0$, then Z is not included in (2.14). 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

$$(2.15) \quad g(x, z) = G(x, z, \theta)$$

for some unknown $\theta \in \Theta$, known function G , and almost every $(x, z) \in [0, 1]^{p_x+r}$. The alternative hypothesis, H_1 is that there is no $\theta \in \Theta$ such that (2.15) holds for almost every $(x, z) \in [0, 1]^{p_x+r}$. The data, $\{Y_i, X_i, Z_i, W_i : i = 1, \dots, n\}$, are a simple random sample of (Y, X, Z, W) .

To define the multivariate extension of τ_n , let f_{YXZW} and f_{ZW} , respectively, denote the probability density function of (Y, X, Z, W) and (Z, W) . Define

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

Let $l_{MV}(z, w; \zeta, \eta)$ be the kernel of a nonsingular operator L_{MV} on $L_2[0, 1]^{p_w+r}$. That is, for any function $\psi \in L_2[0, 1]^{p_w+r}$,

$$L_{MV}(z, w) = \int_{[0,1]^{p_w}} \int_{[0,1]^r} l_{MV}(z, w; \zeta, \eta) d\zeta d\eta.$$

Then H_0 is equivalent to

$$\tau_{MV} = \int_{[0,1]^{p_w}} \int_{[0,1]^r} S_{MV}(z, w)^2 dz dw$$

where

$$S_{MV}(z, w) = \int_{[0,1]^{p_w}} \int_{[0,1]^{p_x}} \int_{[0,1]^r} \{F_{YXZW}[G(x, \zeta, \theta), x, \zeta, \eta] - qf_{ZW}(\zeta, \eta)\} l_{MV}(z, w; \zeta, \eta) d\zeta dx d\eta.$$

Let $S_{n,MV}$ be the following sample analog of S_{MV} :

$$S_{n,MV}(z, w) = n^{-1/2} \sum_{i=1}^n \left\{ I \left[Y_i \leq G(X_i, Z_i, \hat{\theta}) \right] - q \right\} \hat{l}_{MV}(z, w; Z_i, W_i),$$

where $\hat{\theta}$ is an estimator of θ that is consistent under H_0 and \hat{l}_{MV} is a consistent estimator of l_{MV} . The test statistic is

$$\tau_{n,MV} = \int_{[0,1]^{p_w}} \int_{[0,1]^r} S_{n,MV}(z, w)^2 dz dw$$

To obtain the asymptotic distribution of $\tau_{n,MV}$ under H_0 , assume that $\hat{\theta} \rightarrow_p \theta_0$ as $n \rightarrow \infty$ and that

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

for some function γ_{MV} taking values in \mathbb{R}^d such that $E\gamma_{MV}(Y, X, Z, W, \theta_0) = 0$ and $\text{Var}[\gamma_{MV}(Y, X, Z, W, \theta_0)]$ is a finite, non-singular matrix. Set $G_\theta = \partial G / \partial \theta$ and

$$\Gamma_{MV}(z, w) = \int_{[0,1]^{pw}} \int_{[0,1]^{px}} \int_{[0,1]^r} f_{YZW}[G(x, \zeta, \theta_0), x, \zeta, \eta] l_{MV}(z, w; \zeta, \eta) G_\theta(x, \zeta, \theta_0) d\zeta dx d\eta.$$

Define

$$B_{n,MV}(z, w) = n^{-1/2} \sum_{i=1}^n \left[\{I[Y_i \leq g(X_i, Z_i)] - q\} l_{MV}(z, w; Z_i, W_i) + \Gamma_{MV}(z, w)' \gamma_{MV}(Y_i, X_i, Z_i, W_i, \theta_0) \right]$$

and

$$V_{MV}(z_1, w_1; z_2, w_2) = E [B_{n,MV}(z_1, w_1) B_{n,MV}(z_2, w_2)].$$

Let Ω_{MV} denote the operator that is defined by

$$(\Omega_{MV}\psi)(z, w) = \int_{[0,1]^{pw}} \int_{[0,1]^r} V_{MV}(z, w; \zeta, \eta) \psi(\zeta, \eta) d\zeta d\eta.$$

Let $\{\omega_{j,MV} : j = 1, 2, \dots\}$ denote the eigenvalues of Ω_{MV} sorted in decreasing order. Then arguments similar to those used to prove Theorem 1 show that under the regularity conditions given in Section 5.2,

$$\tau_{n,MV} \rightarrow_d \sum_{j=1}^{\infty} \omega_{j,MV} \chi_{1j}^2.$$

One notable regularity condition in Section 5.2 is that the smoothness assumption on the weight function becomes more stringent as $(p+r)$ increases ($\alpha > (p+r)/2$; see Assumption 8 in Section 5.2).

In addition, results analogous to Theorems 3-5 hold for the multivariate statistic. Specifically, the $\tau_{n,MV}$ test:

1. Is consistent against all identified, fixed alternative models;

2. Has power exceeding its level against local alternative models whose distance from the null hypothesis model is $O(n^{-1/2})$;
3. Has asymptotic power exceeding its level uniformly over a class of alternatives whose distance from H_0 is $O(n^{-1/2})$.

An approximate critical value for $\tau_{n,MV}$ can be obtained by modifying the methods of Section 2.4. As in Section 2.4, approximate the asymptotic distribution of $\tau_{n,MV}$ under H_0 by the distribution of

$$\sum_{j=1}^{K_\varepsilon} \omega_{j,MV} \chi_{1j}^2$$

for some constant $K_\varepsilon < \infty$. Let \mathcal{W}_i be a d -vector of instruments (possibly powers of components of Z and W), and let $\hat{\theta}$ satisfy

$$n^{-1/2} \sum_{i=1}^n \mathcal{W}_i \left[I \left\{ Y_i \leq G(X_i, Z_i, \hat{\theta}) \right\} - q \right] = o_p(1).$$

Let $\hat{Y}_i = Y_i - \hat{Q}[Y - G(X, Z, \hat{\theta})|W]$. Let $\hat{L}_{MV}(z, w)$ denote the $n \times 1$ vector whose i th component is $\hat{l}_{MV}(z, w; Z_i, W_i)$, let \hat{F}_{MV} denote the $n \times d$ matrix whose i th row is

$$\frac{1}{\delta_n} K \left(\frac{\hat{Y}_i - G(X_i, Z_i, \hat{\theta})}{\delta_n} \right) G_\theta(X_i, Z_i, \hat{\theta})',$$

and let $\tilde{\mathcal{W}}_{MV}$ denote $n \times d$ matrix whose i th row is \mathcal{W}_i' . Define

$$\hat{\Gamma}_{MV}(z, w) = (n\delta_n)^{-1} \sum_{i=1}^n K \left(\frac{\hat{Y}_i - G(X_i, Z_i, \hat{\theta})}{\delta_n} \right) \hat{l}_{MV}(z, w; Z_i, W_i) G_\theta(X_i, Z_i, \hat{\theta})'$$

and

$$\hat{\Phi}_{MV} = (n\delta_n)^{-1} \sum_{i=1}^n K \left(\frac{\hat{Y}_i - G(X_i, Z_i, \hat{\theta})}{\delta_n} \right) \mathcal{W}_i G_\theta(X_i, Z_i, \hat{\theta})'.$$

Let $\hat{M}_{MV} = I_n - n^{-1} \hat{F}_{MV} \hat{\Phi}_{MV}^{-1} \tilde{\mathcal{W}}_{MV}'$ and

$$\hat{V}_{MV}(z_1, w_1; z_2, w_2) = q(1-q)n^{-1} \hat{L}_{MV}(z_1, w_1)' \hat{M}_{MV} \hat{M}_{MV}' \hat{L}(z_2, w_2).$$

Let $\{\psi_j : j = 1, 2, \dots\}$ be an orthonormal basis for $L_2[0, 1]^{p_w+r}$. Approximate $\hat{l}_{MV}(z, w; \zeta, \eta)$ by the finite sum

$$\Pi_{MV}(z, w; \zeta, \eta) = \sum_{j=1}^J \sum_{k=1}^J \hat{d}_{jk} \psi_j(z, w) \psi_k(\zeta, \eta)$$

for some finite integer $J \geq K_\varepsilon$, where

$$\hat{d}_{jk} = \int_{[0,1]^{pw}} \int_{[0,1]^r} \int_{[0,1]^{pw}} \int_{[0,1]^r} \hat{l}_{MV}(z_1, w_1; z_2, w_2) \psi_j(z_1, w_1) \psi_k(z_2, w_2) dz_1 dw_1 dz_2 dw_2.$$

Let $\psi(z, w)$ denote the $J \times 1$ vector whose j th component is $\psi_j(z, w)$, and let Ψ_{MV} denote the $J \times n$ matrix whose (j, k) th component is $\psi_j(Z_k, W_k)$. Let \hat{D}_{MV} be the $J \times J$ matrix $\{\hat{d}_{jk}\}$. Let $\{\hat{\omega}_{j,MV} : j = 1, \dots, J\}$ be the eigenvalues of the matrix

$$q(1-q)n^{-1} \hat{D}_{MV} \Psi_{MV} \hat{M}_{MV} \hat{M}'_{MV} \Psi'_{MV} \hat{D}'_{MV}.$$

Then the estimated approximate $1 - \alpha$ critical value of $\tau_{n,MV}$ is the $1 - \alpha$ quantile of the distribution of

$$\sum_{j=1}^{K_\varepsilon} \hat{\omega}_{j,MV} \chi_{1j}^2.$$

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, $q = 0.5$. The experiments consist of testing the null hypothesis, H_0 , that

$$(3.1) \quad g(x) = \theta_0 + \theta_1 x$$

against the alternative hypotheses

$$(3.2) \quad g(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

and

$$(3.3) \quad g(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3.$$

In all experiments, $\theta_0 = 0$ and $\theta_1 = 0.5$. When (3.2) is the correct model, $\theta_2 = -0.5$. When (3.3) is the correct model, $\theta_2 = -1.5$ and $\theta_3 = 1$. 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 standard normal distribution function, $\zeta \sim \mathbf{N}(0, 1)$, $\xi = \rho\zeta + \sqrt{1 - \rho^2}\varepsilon$, $\varepsilon \sim \mathbf{N}(0, 1)$, and $\rho = 0.8$. Realizations of Y were generated from $Y = g(x) + \sigma_U U$, where $U = \eta\varepsilon + \sqrt{1 - \eta^2}\nu$, $\nu \sim \mathbf{N}(0, 1)$, $\sigma_U = 0.1$, and $\eta = 0.5$. The τ_n test is obtained with the weight function $l(z, w) = f_{YXW}[g(z), z, w]$ that is estimated

using the method described in Section 2.9. Under H_0 , θ is estimated by the IV quantile-regression estimator of Chernozhukov and Hansen (2006). In particular, θ is estimated with constant weights and instruments $(1, W)$.

To provide a basis for judging whether the τ_n performs well, this section also reports the results of an asymptotic t test of the hypothesis $\theta_2 = 0$. Specifically, θ_2 is estimated by the IV quantile-regression estimator of Chernozhukov and Hansen (2006) under the model (3.2). Constant weights are used and instruments are $(1, W, W^2)$. The t test is an example of an ad hoc test that might be used in applied research. The experiments use sample sizes of $n = 200, 400$ and nominal levels of $\alpha = 0.10, 0.05, 0.01$. There are 500 Monte Carlo replications in each experiment.

To implement the τ_n test, it is necessary to specify tuning parameters. In all procedures involving the kernel function, $K(u) = (15/16)(1 - u^2)I(|u| \leq 1)$ is used. To compute the asymptotic critical value, the basis functions $\{\psi_j(x) = \sqrt{2} \sin(j\pi x) : j = 1, \dots, J\}$ are used. The generalized Fourier coefficients \hat{d}_{jk} in (2.11) were computed by numerical integration (the composite trapezoid rule). The asymptotic critical value was estimated with $J = 10$ and $K_\varepsilon = 10$. Note that J and K_ε are relatively small, because only first a few eigenvalues are important and also higher-order terms of \hat{d}_{jk} are difficult to estimate by numerical integration for a large value of J . In applications, J and K can be chosen large enough that further increases in their values do not affect the value of the test statistic.

The smoothed version of Horowitz's and Lee's (2007) estimator of g can be obtained with the regularization parameter a_n , a bandwidth $h_{n,YXW}$ that is needed to compute the kernel density estimator of f_{YXW} , and another bandwidth h_n that is needed to carry out the smoothing procedure in Section 2.9. To obtain the critical value under sampling from (2.6), \hat{Y}_i 's were generated using a local linear quantile regression estimator of Chaudhuri (1991) with a bandwidth $h_{n,LIQ}$. Also, a bandwidth δ_n is needed to compute $\hat{\Gamma}(z)$ and $\hat{\Phi}$. We have used the following simple, rule-of-thumb, data-driven methods for choosing these tuning parameters. First, $h_{n,YXW}$ is chosen based on the normal reference rule (an extension of Silverman's rule to the multivariate density estimation): $h_{n,YXW} = 2.7054 \times n^{-1/7}$ after re-scaling variables by their standard deviations. The bandwidth $h_{n,LIQ}$ is chosen by a simple rule of thumb suggested by Fan and Gijbels (1996, p.202) and the bandwidth δ_n is chosen by applying Silverman's rule: $\delta_n = 2.7779 \times \hat{\sigma}_U \times n^{-1/5}$, where $\hat{\sigma}_U$ is the sample standard deviation of $\hat{U}_i \equiv Y_i - G(X_i, \hat{\theta})$. The regularization parameter a_n was chosen by $a_n = \sqrt{h_{n,YXW}}$ and this choice ensures that Horowitz's and Lee's (2007) estimator of g

is consistent. The choice of a_n does not make a first-order difference, provided that the estimator of $f_{YXW}[g(z), z, w]$ is consistent. Finally, h_n is chosen by $h_n = 0.5/\log(n)$. One requirement for h_n is that $h_n^{-2}\|\hat{g} - g\| = o_p(1)$. Since \hat{g} can converge very slowly, we have used a sequence that converges to zero at a logarithmic rate. The constant 0.5 is multiplied to $[\log(n)]^{-1}$ since the half the length of the support of X is the maximum possible value for h_n . To check the sensitivity to these choices of bandwidths, we carried out each experiment with two sets of alternative bandwidths: one that are larger by 20% and the other that are smaller by 20%.

The results are shown in Table 1. When H_0 is true, the differences between the nominal and empirical rejection probabilities are small. The results are not sensitive to the bandwidths choice. When H_0 is false and the correct model is (3.2), the power of the τ_n at the 5% level is 0.620 ($n = 200$) and 0.810 ($n = 400$) with the rule-of-thumb bandwidths. The t test is more powerful than the τ_n test. This result is not surprising given that the t test is a consistent test under the alternative model (3.2). However, when H_0 is false and the correct model is (3.3), the t test has no power for $n = 200, 400$. On the other hand, the power of τ_n test is almost 1 for $n = 200$. The τ_n test is shown theoretically to be a consistent test against a general alternative and therefore the results of the Monte Carlo experiments are in line with the theoretical property of the τ_n test.

4 Conclusions

This paper has presented a test of a parametric model of a quantile regression model with a possibly endogenous right-hand side variable against a nonparametric alternative. The model is identified through an instrumental variable. A parametric model typically can be estimated with an $n^{-1/2}$ rate of convergence in probability, whereas nonparametric IV estimators can have much slower rates of convergence. This makes parametric estimation attractive for applied research provided that there is justification for believing that the parametric model is free of serious specification errors. This paper provides a specification test. Under mild conditions, the test is consistent against any alternative model. In addition, in large samples, the test's power is arbitrarily close to 1 uniformly over a class of alternative models whose distance from the parametric model is $O(n^{-1/2})$. Some Monte Carlo experiments have illustrated the satisfactory finite-sample performance of the test.

5 Appendix

5.1 Proofs of Theorems

The asymptotic distribution of τ_n can be obtained using arguments similar to those used in Horowitz (2006), combined with empirical process methods of Van der Vaart and Wellner (1996).

Let $\{\mathbf{X}_i \equiv (Y_i, X_i, W_i) : i = 1, 2, \dots, n\}$ denote observed data with sample size n . Let \mathcal{H} be a class of measurable functions with a measurable envelope function H . Let $N(\varepsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})$ and $N_{[\cdot]}(\varepsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})$, respectively, denote the covering and bracketing numbers for the set \mathcal{H} (for exact definitions, see, for example, Van der Vaart and Wellner (1996, p.83)). In addition, let $J_{[\cdot]}(1, \mathcal{H}, L^2(P))$ denote a bracketing integral of \mathcal{H} , that is

$$J_{[\cdot]}(1, \mathcal{H}, L^2(P)) = \int_0^1 \sqrt{1 + \log N_{[\cdot]}(\varepsilon \|H\|_{L^2(P)}, \mathcal{H}, L^2(P))} d\varepsilon,$$

where $\|\cdot\|_{L^2(P)}$ is the L^2 -norm with probability measure P . We will use the following lemma, which is due to the last display of Theorem 2.14.2 of Van der Vaart and Wellner (1996, p.240) and has also been used in Ichimura and Lee (2006).

Lemma 5.1. *Assume that $\{\mathbf{X}_i : i = 1, 2, \dots, n\}$ is a random sample of \mathbf{X} . Let \mathcal{H} be a class of measurable functions with a measurable envelope function H . Then there exists a constant C such that*

$$E \left[\sup_{h \in \mathcal{H}} \left| n^{-1/2} \sum_{i=1}^n \{h(\mathbf{X}_i) - E[h(\mathbf{X})]\} \right| \right] \leq C J_{[\cdot]}(1, \mathcal{H}, L^2(P)) \|H\|_{L^2(P)}.$$

To obtain the asymptotic distribution of τ_n , rewrite $S_n(z)$ as

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

where

$$\begin{aligned} S_{n1}(z) &= n^{-1/2} \sum_{i=1}^n [I\{Y_i \leq g(X_i)\} - q] l(z, W_i), \\ S_{n2}(z) &= n^{-1/2} \sum_{i=1}^n [I\{Y_i \leq G(X_i, \theta_0)\} - I\{Y_i \leq g(X_i)\}] l(z, W_i), \\ S_{n3}(z) &= n^{-1/2} \sum_{i=1}^n [I\{Y_i \leq G(X_i, \hat{\theta})\} - I\{Y_i \leq G(X_i, \theta_0)\}] l(z, W_i), \end{aligned}$$

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

$$S_{n5}(z) = n^{-1/2} \sum_{i=1}^n [I \{Y_i \leq G(X_i, \theta_0)\} - I \{Y_i \leq g(X_i)\}] [\hat{l}(z, W_i) - l(z, W_i)],$$

and

$$S_{n6}(z) = n^{-1/2} \sum_{i=1}^n [I \{Y_i \leq G(X_i, \hat{\theta})\} - I \{Y_i \leq G(X_i, \theta_0)\}] [\hat{l}(z, W_i) - l(z, W_i)].$$

Lemma 5.2. *As $n \rightarrow \infty$,*

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

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

Proof. Define

$$S_3(\mathbf{X}_i; \theta, z) = [I \{Y_i \leq G(X_i, \theta)\} - I \{Y_i \leq G(X_i, \theta_0)\}] l(z, W_i)$$

and

$$S_3^*(\theta, z) = E \left\{ [I \{Y \leq G(X, \theta)\} - I \{Y \leq G(X, \theta_0)\}] l(z, W) \right\}.$$

Note that by a Taylor series expansion,

$$(5.1) \quad S_3^*(\hat{\theta}, z) = \Gamma(z)' (\hat{\theta} - \theta_0) + O \left[(\hat{\theta} - \theta_0)^2 \right]$$

uniformly over z .

For any ε , define $\mathcal{N}_3(\varepsilon) = \{(\theta, z) : \|\theta - \theta_0\| \leq \varepsilon \text{ and } z \in [0, 1]\}$. Observe that using arguments similar to those used in Chen, Linton and Van Keilegom (2003, pp.1599-1600),

$$\begin{aligned} \sup_{(\theta, z) \in \mathcal{N}_3(\varepsilon)} |S_3(\mathbf{X}_i; \theta, z)| &\leq [I \{Y_i \leq G(X_i, \theta_0) + \varepsilon C_g\} - I \{Y_i \leq G(X_i, \theta_0) - \varepsilon C_g\}] C_l \\ &\equiv \bar{S}_3(\mathbf{X}_i). \end{aligned}$$

Thus, $\bar{S}_3(\mathbf{x})$ is an envelope function for the class $\mathcal{S}_3(\varepsilon) \equiv \{S_3(\mathbf{x}; \theta, z) : (\theta, z) \in \mathcal{N}_3(\varepsilon)\}$ and $\|\bar{S}_3\|_{L^2(P)} \leq C\varepsilon^{1/2}$ for some finite constant C . Since $G(x, \theta)$ is finite-dimensional and $l(z, w)$ is Lipschitz continuous with respect to z uniformly over w , it can be shown that $J_{[\cdot]}(1, \mathcal{S}_3(\varepsilon), L^2(P)) < \infty$ for any ε . Then by Lemma 5.1,

$$(5.2) \quad \sup_{(\theta, z) \in \mathcal{N}_3(n^{-1/2})} \left| n^{-1/2} \sum_{i=1}^n [S_3(\mathbf{X}_i; \theta, z) - S_3^*(\theta, z)] \right| \leq Cn^{-1/4}.$$

Therefore, the lemma follows immediately from (5.1) and (5.2). \square

Lemma 5.3. As $n \rightarrow \infty$,

$$S_{n4}(z) = o_p(1)$$

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

Proof. This can be proved again by the empirical process technique used in the proof of Lemma 5.2. To do so, define, for any function $\delta(z, w)$,

$$S_4(\mathbf{X}_i; \delta, z) = [I \{Y_i \leq g(X_i)\} - q] \delta(z, W_i)$$

and

$$S_4^*(\delta, z) = E \left\{ [I \{Y \leq g(X)\} - q] \delta(z, W) \right\}.$$

Obviously, $S_4^*(\delta, z) \equiv 0$ since W is an instrument. Given that $\hat{l}(z, w)$ is uniformly consistent for $l(z, w)$, define $\mathcal{N}_4(\varepsilon) = \{(\delta, z) : \sup_{(z,w) \in [0,1]^2} |\delta(z, w)| \leq \varepsilon \text{ and } z \in [0, 1]\}$ for any ε . Consider the class $\mathcal{S}_4(\varepsilon) \equiv \{S_4(\mathbf{x}; \delta, z) : (\delta, z) \in \mathcal{N}_4(\varepsilon)\}$ with an envelope function ε . Since $l(z, w)$ and $\hat{l}(z, w)$ are uniformly Lipschitz continuous with respect to both z and w , it can be proved that

$$(5.3) \quad J_{[\cdot]}(1, \mathcal{S}_4(\varepsilon), L^2(P)) < \infty$$

by arguments identical to those used in the calculation of covering numbers in the proof of Lemmas B.2 and B.3 of Ichimura and Lee (2006). Then the lemma follows from Lemma 5.1. \square

Lemma 5.4. As $n \rightarrow \infty$,

$$S_{n6}(z) = o_p(1)$$

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

Proof. This can be proved again by arguments similar to those used in the proofs of Lemmas 5.2 and 5.3. \square

Proof of Theorem 1. Under H_0 , by Lemmas 5.2, 5.3, and 5.4,

$$S_n(z) = B_n(z) + o_p(1)$$

uniformly over $z \in [0, 1]$. Then the theorem follows by arguments identical to those used in the proof of Theorem 1 of Horowitz (2006, Supplement). \square

Proof of Theorem 2. Using arguments similar to those used in the proof of Lemma A.9 of Lee (2007), it can be shown that $\sup_{0 \leq z \leq 1} |\hat{\Gamma}(z) - \Gamma(z)| = o_p(1)$ and $\|\hat{\Phi} - \Phi\| = o_p(1)$. By Theorem 5.1a of Bhatia, Davis, and McIntosh (1983), $|\hat{\omega}_j - \tilde{\omega}_j| = O\left(\|\hat{\Omega} - \tilde{\Omega}\|\right)$. Then Part (i) of the theorem follows by the assumption that $\hat{\Gamma}(z)$ and $\hat{\Phi}$ are consistent estimators. Part (ii) follows immediately from part (i). \square

Proof of Theorem 3. As in the proof of Theorem 3 of Horowitz (2006, Supplement), it can be shown that

$$n^{-1/2}S_n(z) \rightarrow_p \mathcal{T}(z)$$

uniformly over $z \in [0, 1]$. Then the theorem follows immediately. \square

Proof of Theorem 4. If $\hat{\theta}$ satisfies (2.7), then under the sequence of local alternatives (2.12), it can be shown that

$$\begin{aligned} n^{1/2}(\hat{\theta} - \theta_0) &= -n^{-1/2} \sum_{i=1}^n [\tilde{\gamma}(Y_i, X_i, W_i, \theta_0) - E\tilde{\gamma}(Y, X, W, \theta_0)] \\ &\quad + \Phi^{-1}E[f_{U|X,W}(0|X, W)\Delta(X)\mathcal{W}] + o_p(1). \end{aligned}$$

Furthermore, using arguments similar to those used to prove Lemmas 5.2 and 5.3, it can be shown that under the sequence of local alternatives (2.12), as $n \rightarrow \infty$,

$$S_{n2}(z) = -E[f_{U|X,W}(0|X, W)\Delta(X)l(z, W)] + o_p(1)$$

and

$$S_{n5}(z) = o_p(1)$$

uniformly over $z \in [0, 1]$. Then under the sequence of local alternatives (2.12),

$$S_n(z) = B_n(z) + \mu(z) + o_p(1)$$

uniformly over $z \in [0, 1]$. Then the theorem follows by arguments identical to those used in the proof of Theorem 4 of Horowitz (2006, Supplement). \square

Proof of Theorem 5. The proof here is similar to that of Theorem 5 of Horowitz (2006, Supplement). Thus, instead of following all the steps in the proof of Theorem 5 of Horowitz (2006, Supplement), we sketch the proof and point out the main differences.

Define

$$S_2^*(g, z) = E\left\{ [I\{Y \leq G(X, \theta_0)\} - I\{Y \leq g(X)\}] l(z, W) \right\}$$

and

$$S_5^*(g, \delta, z) = E\left\{ [I\{Y \leq G(X, \theta_0)\} - I\{Y \leq g(X)\}] \delta(z, W) \right\}$$

for any fixed function $\delta(z, w)$. Further, define $D_n(z) = S_{n3}(z) + S_{n6}(z) + n^{1/2}S_2^*(g, z) + n^{1/2}S_5^*(g, \hat{l} - l, z)$ and $\tilde{S}_n(z) = S_n(z) - D_n(z)$. Then using the empirical process method combined with the assumption that $g \in \mathcal{C}_{C_g}^\alpha([0, 1])$ with $\alpha > 1$ and some finite constant C_g , it can be shown that $\tilde{S}_n(z)$ is bounded in probability uniformly over $g \in \mathcal{F}_{nc}$ and $z \in [0, 1]$. This in turn implies that $\|\tilde{S}_n\|$ is bounded in probability uniformly over $g \in \mathcal{F}_{nc}$.

By arguments identical to those used in the proof of Theorem 5 of Horowitz (2006, Supplement), for each $\varepsilon > 0$, there is M_ε such that, for all $M > M_\varepsilon$,

$$\Pr(\tau_n > \tilde{z}_\alpha) \geq \Pr(0.5 \|D_n\|^2 > \tilde{z}_\alpha + M) - \varepsilon.$$

Now notice that

$$S_5^*(g, \hat{l} - l, z) = O\left(\|Q_g\| \|\hat{l} - l\|\right).$$

This implies that under the restriction that $\sup_{g \in \mathcal{F}_{nc}} \|Q_g\| \|\hat{l} - l\| / \|\mathcal{T}_g\| = o_p(1)$,

$$n^{1/2}S_2^*(g, z) + n^{1/2}S_5^*(g, \hat{l} - l, z) \geq 0.5n^{1/2}(\mathcal{T}_g)(z)$$

uniformly over $g \in \mathcal{F}_{nc}$ for all sufficiently large n . In addition, using empirical process arguments again gives

$$\|S_{n3} + S_{n6}\| = O_p(1)$$

uniformly over $g \in \mathcal{F}_{nc}$. Then the remaining part of the proof can be completed by repeating the arguments in the proof of Theorem 5 of Horowitz (2006, Supplement, page 7). \square

Proof of Claims in Section 2.9. To show claim (iii) in Section 2.9, note that if $h_n \leq z \leq (1 - h_n)$, then

$$\tilde{g}(z) - g(z) = \int_{-1}^1 K(\xi) [g(z + h_n\xi) - g(z)] d\xi$$

and

$$(5.4) \quad \begin{aligned} |\tilde{g}(z) - g(z)| &\leq \int_{-1}^1 K(\xi) |g(z + h_n \xi) - g(z)| d\xi \\ &\leq C_g h_n \end{aligned}$$

for some constant $C_g < \infty$. Now define

$$\tilde{\mu}_g(z) = - \int_0^1 \int_0^1 f_{YXW}(g(x), x, w) \Delta(x) f_{YXW}(\tilde{g}(z), z, w) dx dw.$$

Then it follows from (5.4) that

$$\|\bar{\mu}_g\| - C_3 h_n \leq \|\tilde{\mu}_g\| \leq \|\bar{\mu}_g\| - C_3 h_n$$

for some constant $C_3 < \infty$. Moreover,

$$\begin{aligned} (C_l C_\Delta)^{1/2} \|\bar{\mu}_g\| &\leq (C_l C_\Delta)^{1/2} \|\tilde{\mu}_g\| \left(1 + \frac{C_3 h_n}{\|\tilde{\mu}_g\|}\right) \\ &\leq (C_l C_\Delta)^{1/2} \|\tilde{\mu}_g\| \left(1 + \frac{C_3 h_n}{\|\bar{\mu}_g\| - C_3 h_n}\right). \end{aligned}$$

Therefore, (2.13) holds if g is replaced by \tilde{g} and h_n is sufficiently small.

We now consider (i). It is easy to see that if g is Lipschitz continuous, then so is $\tilde{g}(z)$ on $h_n \leq z \leq 1 - h_n$. Also,

$$\begin{aligned} \tilde{g}(z_2) - \tilde{g}(z_1) &= h_n^{-1} \int_0^1 \left[K\left(\frac{x - z_2}{h_n}\right) - K\left(\frac{x - z_1}{h_n}\right) \right] \hat{g}(x) dx \\ &= h_n^{-1} \int_0^1 \left[K\left(\frac{x - z_2}{h_n}\right) - K\left(\frac{x - z_1}{h_n}\right) \right] g(x) dx \\ &\quad + h_n^{-1} \int_0^1 \left[K\left(\frac{x - z_2}{h_n}\right) - K\left(\frac{x - z_1}{h_n}\right) \right] [\hat{g}(x) - g(x)] dx \\ &\equiv B_1(z_2, z_1) + B_2(z_2, z_1). \end{aligned}$$

Assume that $h_n \leq z_2, z_1 \leq 1 - h_n$. Then

$$B_1(z_2, z_1) = \int_{-1}^1 K(\xi) [g(h_n \xi + z_2) - g(h_n \xi + z_1)] d\xi.$$

It follows that

$$|B_1(z_2, z_1)| \leq C_g |z_2 - z_1|.$$

In addition, for some constant $C_K < \infty$,

$$\begin{aligned} |B_2(z_2, z_1)| &\leq \frac{C_K}{h_n^2} |z_2 - z_1| \int_0^1 |\hat{g}(x) - g(x)| dx \\ &\leq C_K \frac{\|\hat{g} - g\|}{h_n^2} |z_2 - z_1|. \end{aligned}$$

Therefore, it follows from the triangle inequality that

$$|\tilde{\hat{g}}(z_2) - \tilde{\hat{g}}(z_1)| \leq \left(C_g + C_K \frac{\|\hat{g} - g\|}{h_n^2} \right) |z_2 - z_1|.$$

Hence, $\tilde{\hat{g}}(z)$ is Lipschitz continuous on $h_n \leq z \leq 1 - h_n$ with probability approaching 1 since $h_n^{-2} \|\hat{g} - g\| = o_p(1)$.

To show the uniform convergence of $\tilde{\hat{g}}(z)$ to $\tilde{g}(z)$, note that

$$\tilde{\hat{g}}(z) - \tilde{g}(z) = h_n^{-1} \int_0^1 K\left(\frac{x-z}{h_n}\right) [\hat{g}(x) - g(x)] dx.$$

Then, for any $h_n \leq z \leq 1 - h_n$,

$$\begin{aligned} |\tilde{\hat{g}}(z) - \tilde{g}(z)| &= \frac{C_K}{h_n} \int_0^1 |\hat{g}(x) - g(x)| dx \\ &\leq C_K \frac{\|\hat{g} - g\|}{h_n}. \end{aligned}$$

Therefore, it follows from $h_n^{-2} \|\hat{g} - g\| = o_p(1)$ that

$$\sup_{h_n \leq z \leq 1 - h_n} |\tilde{\hat{g}}(z) - \tilde{g}(z)| = o_p(1).$$

Then (i) follows from the proofs of Theorems 1 and 3 and the assumption that $h_n \rightarrow \infty$ as $n \rightarrow \infty$. Finally, (ii) can be proved using the arguments identical to those used in the proof of Theorem 5 with the restriction that $h_n \leq z \leq 1 - h_n$. \square

5.2 Regularity Conditions for the Multivariate Extension

This section states the assumptions that are used to obtain the asymptotic properties of $\tau_{n,MV}$ under the null and alternative hypotheses.

Assumption 5. (i) The support of (X, Z, W) is $[0, 1]^{p_x+r+p_w}$. (ii) (Y, X, Z, W) has a probability density function f_{YXZW} with respect to Lebesgue measure. (iii) f_{YXZW} is bounded and is differentiable with respect to its first argument with a uniformly bounded derivative. (iv) $\{(Y_i, X_i, Z_i, W_i) : i = 1, \dots, n\}$ is a simple random sample of (Y, X, Z, W) .

Assumption 6. *There is a constant $C_{G,MV} < \infty$ such that $\sup_{(x,z) \in [0,1]^{p_x+r}} |g(x,z)| \leq C_{G,MV}$, $\sup_{\theta \in \Theta} \sup_{(x,z) \in [0,1]^{p_x+r}} |G(x,z,\theta)| \leq C_{G,MV}$, and the first and second derivatives of $G(x,z,\theta)$ with respect to θ are bounded by $C_{G,MV}$ uniformly over $(x,z) \in [0,1]^{p_x+r}$ and $\theta \in \Theta$.*

Assumption 7. *(i) As $n \rightarrow \infty$, $\hat{\theta} \rightarrow_p \theta_0$ for some $\theta_0 \in \Theta$, a compact subset of \mathbb{R}^d . (ii) If H_0 is true, then $g(x,z) \equiv G(x,z,\theta_0)$, $\theta_0 \in \text{int}(\Theta)$, and*

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

for some function γ_{MV} taking values in \mathbb{R}^d such that $E[\gamma_{MV}(Y, X, Z, W, \theta_0)] = 0$ and $\text{Var}[\gamma_{MV}(Y, X, Z, W, \theta_0)]$ is a finite, nonsingular matrix.

Assumption 8. *(i) $\sup_{(z,w;\zeta,\eta) \in [0,1]^{2(p_w+r)}} |\hat{l}(z,w;\zeta,\eta) - l(z,w;\zeta,\eta)| = o_p(1)$. (ii) The operator L_{MV} is nonsingular. (iii) There is a constant $C_{l,MV} < \infty$ such that*

$$\begin{aligned} \sup_{(z,w;\zeta,\eta) \in [0,1]^{2(p_w+r)}} |l_{MV}(z,w;\zeta,\eta)| &\leq C_{l,MV}, \\ \sup_{(\zeta,\eta) \in [0,1]^{p_w+r}} |l_{MV}(z_1,w_1;\zeta,\eta) - l_{MV}(z_2,w_2;\zeta,\eta)| &\leq C_{l,MV} \|(z_1,w_1) - (z_2,w_2)\|, \end{aligned}$$

and for each fixed (z,w) , the class of functions $\{l_{MV}(z,w;\zeta,\eta)\}$ belongs to $\mathcal{C}_{C_{l,MV}}^\alpha([0,1]^{p_w+r})$ with some $\alpha > (p_w + r)/2$. In addition, with probability approaching 1 as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{(z,w;\zeta,\eta) \in [0,1]^{2(p_w+r)}} |\hat{l}_{MV}(z,w;\zeta,\eta)| &\leq C_{l,MV}, \\ \sup_{(\zeta,\eta) \in [0,1]^{p_w+r}} |\hat{l}_{MV}(z_1,w_1;\zeta,\eta) - \hat{l}_{MV}(z_2,w_2;\zeta,\eta)| &\leq C_{l,MV} \|(z_1,w_1) - (z_2,w_2)\|, \end{aligned}$$

and for each fixed (z,w) , the class of functions $\{\hat{l}_{MV}(z,w;\zeta,\eta)\}$ belongs to $\mathcal{C}_{C_{l,MV}}^\alpha([0,1]^{p_w+r})$ with some $\alpha > (p_w + r)/2$.

These assumptions are a straightforward multivariate generalization of regularity condition in Section 2.2. Notice that the smoothness assumption on the weight function in Assumption 8 becomes more stringent as $(p_w + r)$ increases. This condition is needed to prove a multivariate extension of (5.3) using arguments those used in the calculation of covering numbers in the proof of Lemmas B.2 and B.3 of Ichimura and Lee (2006).

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.
- 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 (2007): Semi-nonparametric IV estimation of shape-invariant Engel curves, *Econometrica*, 75, 1613-1669.
- Chaudhuri, P. (1991): Nonparametric estimation of regression quantiles and their local Bahadur representation, *Annals of Statistics*, 19, 760-777.
- Chen, X., O. Linton, and I. Van Keilegom (2003): Estimation of semiparametric models when the criterion function is not smooth, *Econometrica*, 71, 1591-1608.
- 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.
- 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 (2006): 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 estimation of nonseparable models, *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.
- Darolles, S., J. -P. Florens, and E. Renault (2006): Nonparametric instrumental regression, Working paper, GREMAQ, University of Social Science, Toulouse, France.
- Donald, S. G., G. W. Imbens, and W. K. Newey (2003): Empirical likelihood estimation and consistent tests with conditional moment restrictions, *Journal of Econometrics*, 117, 55-93.
- Guerre, E., and P. Lavergne (2002): Optimal minimax rates for nonparametric specification testing in regression models, *Econometric Theory*, 18, 1139-1171.
- Hall, P. and J. L. Howowitz (2005): Nonparametric methods for inference in the presence of instrumental variables, *Annals of Statistics*, 33, 2904-2929.
- Honoré, B.E. and L. Hu (2004): On the performance of some robust instrumental variables estimators, *Journal of Business and Economic Statistics*, 22, 30-39.
- Horowitz, J. L. (2006): Testing a parametric model against a nonparametric alternative with identification through instrumental variables, *Econometrica*, 74, 521-538.
- Horowitz, J. L. and S. Lee (2007): Nonparametric instrumental variables estimation of a quantile regression model, *Econometrica*, 75, 1191-1208.
- 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.
- Ichimura, H. and S. Lee (2006): Characterization of the asymptotic distribution of semi-parametric M-estimators, Cemmap Working Papers, CWP15/06, available at: <http://cemmap.ifs.org.uk>.

- 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.
- Lee, S. (2007): Endogeneity in quantile regression models: a control function approach, *Journal of Econometrics*, 141, 1131-1158.
- Ma, L. and R. Koenker (2006): Quantile regression methods for recursive structural equation models, *Journal of Econometrics*, 134, 471-506.
- 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 models, *Econometrica*, 67, 565-603.
- Powell, J. L. (1983): The asymptotic normality of two-stage least absolute deviations estimators, *Econometrica*, 51, 1569-1576.
- Sakata, S. (2007): Instrumental variable estimation based on conditional median restriction, *Journal of Econometrics*, 141, 350-382.
- Tripathi, G. and Y. Kitamura (2003): Testing conditional moment restrictions, *Annals of Statistics*, 31, 2059-2095.
- Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Table 1. *Results of Monte Carlo Experiments*

Null Model	Alternative Model	Sample Size	Bandwidth	Empirical Probability that H_0 is Rejected Using							
				τ_n test			t test				
				Nominal Level			Nominal Level				
				0.10	0.05	0.01	0.10	0.05	0.01		
H_0 is true											
(3.1)		200	$c = 1.0$	0.094	0.040	0.010	0.088	0.032	0.004		
			$c = 0.8$	0.090	0.036	0.010					
			$c = 1.2$	0.092	0.032	0.008					
		400	$c = 1.0$	0.086	0.050	0.016	0.086	0.042	0.010		
			$c = 0.8$	0.084	0.050	0.014					
			$c = 1.2$	0.086	0.048	0.016					
		H_0 is false									
		(3.1)	(3.2)	200	$c = 1.0$	0.698	0.620	0.454	0.770	0.666	0.470
					$c = 0.8$	0.718	0.628	0.462			
$c = 1.2$	0.670				0.594	0.426					
400	$c = 1.0$			0.864	0.810	0.620	0.962	0.932	0.788		
	$c = 0.8$			0.902	0.846	0.688					
	$c = 1.2$			0.810	0.734	0.556					
(3.1)	(3.3)			200	$c = 1.0$	0.992	0.992	0.992	0.084	0.030	0.004
					$c = 0.8$	0.994	0.992	0.990			
					$c = 1.2$	0.992	0.992	0.992			
		400	$c = 1.0$	1.000	1.000	1.000	0.084	0.050	0.008		
			$c = 0.8$	1.000	1.000	1.000					
			$c = 1.2$	1.000	1.000	1.000					

Notes: Bandwidths with $c = 1.0$ are chosen by some rule-of-thumb, data-driven methods. See descriptions in Section 3. Bandwidths with $c = 0.8$ and $c = 1.2$ are those that are smaller by 20% and larger by 20%, respectively, than the rule-of-thumb bandwidths.