Review of Economic Studies (2007) **74**, 1035–1058 0034-6527/07/00361035$02.00

A Non-Parametric Test of Exogeneity

RICHARD BLUNDELL
University College London
and
JOEL L. HOROWITZ
Northwestern University

First version received February 2005; *final version accepted January* 2007 (*Eds.*)

This paper presents a test for exogeneity of explanatory variables that minimizes the need for auxiliary assumptions that are not required by the definition of exogeneity. It concerns inference about a non-parametric function g that is identified by a conditional moment restriction involving instrumental variables (IV). A test of the hypothesis that g is the mean of a random variable Y conditional on a covariate X is developed that is not subject to the ill-posed inverse problem of non-parametric IV estimation. The test is consistent whenever g differs from $E(Y \mid X)$ on a set of non-zero probability. The usefulness of this new exogeneity test is displayed through Monte Carlo experiments and an application to estimation of non-parametric consumer expansion paths.

1. INTRODUCTION

The problem of endogeneity arises frequently in economics. In empirical microeconomics, endogeneity usually occurs as a result of the joint determination of observed variables by individual agents. For example, firms choose inputs and production levels, and households choose consumption levels and labour supply. It has long been understood that the econometric estimation methods needed when a model contains endogenous explanatory variables are different from those that suffice when all variables are exogenous. For example, ordinary least squares (OLS) does not provide consistent estimates of the coefficients of a linear model when one or more explanatory variables are endogenous. Therefore, it is important to have ways of testing for exogeneity of a model's explanatory variables.

This paper describes an exogeneity test that is applicable in a broad range of circumstances and minimizes the need for auxiliary assumptions that are not required by the definition of exogeneity. The specific case that motivates our research and provides the application in this paper concerns the non-parametric analysis of consumer behaviour and, in particular, the possible endogeneity of total outlay in the non-parametric estimation of Engel curves (expansion paths). Suppose we are interested in estimating the structural relationship between the quantity of leisure services bought by a consumer in a particular month and his total consumption (or wealth) in that month. Knowledge of the shape of this kind of Engel curve is an integral part of any analysis of consumer welfare (Deaton, 1998) and is also a key input into micro-data-based revealed-preference bounds (see Blundell, Browning and Crawford, 2003). It is important to allow this relationship to vary flexibly, which can be done by using non-parametric regression methods. However, it is likely that the unobservables in the relationship, which include individual tastes for leisure, are related to preferences for overall consumption or wealth. If this is the case, then the total consumption (expenditure) variable will be endogenous for the Engel curve, and standard non-parametric regression estimators will not recover the structural relationship of interest for welfare or revealed-preference analysis.[1]

The structural function can be estimated in the presence of endogenous explanatory variables if we have instruments for those variables that are mean independent of the unobservable error term in the structural relationship. Indeed, instrumental variables (IV) estimators for linear models are well known and widely used in empirical economics. When the structural (or regression) function is non-parametric, as in the case considered in this paper, and there is an endogenous explanatory variable, the precision of any estimator is typically much lower than it is when all explanatory variables are exogenous (Hall and Horowitz, 2005). Consequently, there is a large loss of estimation efficiency from unnecessarily treating one or more explanatory variables as endogenous. On the other hand, erroneously assuming exogeneity produces a specification error that may cause the estimation results to be highly misleading. Therefore, it is important to have ways to test for exogeneity in non-parametric regression analysis. This paper presents the first such test.

The approach taken in this paper is to test the orthogonality condition that defines the null hypothesis of exogeneity. In a linear regression model, there are several asymptotically equivalent tests of this condition (Smith, 1994). In non-parametric regression, one possible approach is to compare a non-parametric estimate of the regression function under exogeneity with an estimate obtained by using non-parametric IV methods. Non-parametric IV estimators of the structural function have been developed by Darolles, Florens and Renault (2006), Blundell, Chen and Kristensen (2007), Newey and Powell (2003), and Hall and Horowitz (2005).[2] However, the moment condition that identifies the structural function in the presence of endogeneity is a Fredholm 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 non-parametric estimator of the structural function is typically very slow. Therefore, a test based on a direct comparison of non-parametric estimates obtained with and without assuming exogeneity is likely to have very low power. Accordingly, it is desirable to have a test of exogeneity that avoids non-parametric IV estimation of the structural relationship. This paper presents such a test.

If the structural regression function is known up to a finite-dimensional parameter, then exogeneity can be tested by using methods developed by Hausman (1978), Bierens (1990), and Bierens and Ploberger (1997). However, these tests can give misleading results if the structural function is misspecified. The non-parametric test we present avoids this problem. Another possibility is to test for the exclusion of the instruments from the structural regression. However, we will show that an omitted variables test imposes stronger restrictions than are implied by the hypothesis of exogeneity. It is desirable to avoid these restrictions if exogeneity is the hypothesis of interest. Our test accomplishes this and is no more difficult to implement than an omitted variables test.

Computation of the test statistic and its critical value require only finite-dimensional matrix manipulations, kernel non-parametric regression, and kernel non-parametric density estimation. A GAUSS program for computing the statistic is available at the *Review*'s website.

Section 2 of this paper presents the test. This section also explains the difference between testing for exogeneity and testing for omitted IV in a mean regression. Section 3 describes the asymptotic properties of the test. In Section 4, we present the results of a Monte Carlo investigation of the finite-sample performance of the test. Section 5 presents an application that consists

1. See Blundell and Powell (2003) for a discussion of structural functions of interest in non-parametric regression.

2. Newey, Powell and Vella (1999) developed a non-parametric IV estimator based on "control functions". This estimator avoids the problems described in the remainder of this paragraph, but its assumptions are considerably stronger than ours or those of the authors just cited.

of testing the hypothesis that the income variable in an Engel curve is exogenous. Section 6 concludes. The proofs of theorems are in the appendix.

2. THE MODELLING FRAMEWORK AND THE TEST STATISTIC

This section begins by presenting a detailed description of the model setting that we deal with and the test statistic. Section 2.3 explains why our test is not a test for omitted variables.

2.1. *The model setting*

To be more precise about the setting for our analysis, let Y be a scalar random variable, X and W be continuously distributed random scalars or vectors, and g be a structural function that is identified by the relation

$$E[Y - g(X) \mid W] = 0. \tag{2.1}$$

In (2.1), Y is the dependent variable, X is the explanatory variable, and W is an instrument for X. The function g is non-parametric; it is assumed to satisfy mild regularity conditions but is otherwise unknown.

Define the conditional mean function $G(x) = E(Y \mid X = x)$. We say that X is exogenous if $g(x) = G(x)$ except, possibly, if x is contained in a set of zero probability. Otherwise, we say that X is endogenous. This paper presents a test of the null hypothesis, H_0, that X is exogenous against the alternative hypothesis, H_1, that X is endogenous. It follows from (2.1) that this is equivalent to testing the hypothesis $E[Y - G(X) \mid W] = 0$. Under mild conditions, the test rejects H_0 with probability approaching 1 as the sample size increases whenever $g(x) \neq G(x)$ on a set of non-zero probability.

To understand the issues involved in estimating $g(x)$ when X is endogenous, write (2.1) in the form

$$E(Y \mid W) = \int g(x) dF_{X|W}, \tag{2.2}$$

where $F_{X|W}$ is the cumulative distribution function of X conditional on the instrument W. Equation (2.2) is an integral equation for the structural function g. Identifiability of g is equivalent to uniqueness of the solution of this integral equation. Assuming that that g is identified, estimating it amounts to solving (2.2) after replacing $E(Y \mid W)$ and $F_{X|W}$ with consistent estimators. Doing this is complicated, however, because (2.2) is a version of a Fredholm integral equation of the first kind (O'Sullivan, 1986; Kress, 1999), and it produces a so-called ill-posed inverse problem. Specifically, the solution to (2.2) is not a continuous functional of $E(Y \mid W)$, even if the solution is unique, and $E(Y \mid W)$ and $F_{X|W}$ are smooth functions. Therefore, very different structural functions g can yield very similar reduced forms $E(Y \mid W)$. A similar problem arises in linear regression with multicollinearity, where large differences in regression coefficients can correspond to small differences in the fitted values of the regression function. As a consequence of the ill-posed inverse problem, the rate of convergence in probability of a non-parametric IV estimator of g is typically very slow. Depending on the details of the distribution of (Y, X, W), the rate may be slower than $O_p(n^{-\varepsilon})$ for any $\varepsilon > 0$ (Hall and Horowitz, 2005).

The test developed here does not require non-parametric estimation of g and is not affected by the ill-posed inverse problem of non-parametric IV estimation. Consequently, the "precision" of the test is greater than that of any non-parametric 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, the test can detect a large class of structural functions g whose distance from the conditional mean function G in a suitable metric is $O(n^{-1/2})$. In contrast, the rate of convergence in probability of a non-parametric estimator of g is always slower than $O_p(n^{-1/2})$.[3]

Throughout the remaining discussion, we will use an extended version of (2.1) 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

$$Y = g(X, Z) + U; \quad E(U \mid Z, W) = 0, \tag{2.3}$$

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 (2.3). W is an instrument for X. The inferential problem is to test the null hypothesis, H_0, that

$$\boldsymbol{E}(U \mid X = x, Z = z) = 0, \tag{2.4}$$

except, possibly, if (x, z) belongs to a set of probability 0. The alternative hypothesis, H_1, is that (2.4) does not hold on some set $B \subset [0,1]^{p+r}$ that has non-zero probability. The data, $\{Y_i, X_i, Z_i, W_i : i = 1, \ldots, n\}$ are a simple random sample of (Y, X, Z, W).

2.2. *The test statistic*

To form the test statistic, let f_{XZW} denote the probability density function of (X, Z, W). Define $G(x,z) = \boldsymbol{E}(Y \mid X = x, Z = z)$. In what follows, we use operator notation that is taken from functional analysis and is widely used in the literature on non-parametric IV estimation. See, for example, Darolles *et al.* (2002), Carrasco, Florens and Renault (2005), Hall and Horowitz (2005), and Horowitz (2006). For each $z \in [0,1]^r$, define the operator T_z on $L_2[0,1]^p$ by

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

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

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

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

$$\tilde{S}(x,z) \equiv T_z(g - G)(x,z) = 0, \tag{2.5}$$

for almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that (2.5) does not hold on a set $B \subset [0,1]^{p+r}$ with non-zero Lebesgue measure. A test statistic can be based on a sample analogue of $\int \tilde{S}(x,z)^2 dxdz$, but the resulting rate of testing is slower than $n^{-1/2}$ if $r > 0$. The rate $n^{-1/2}$ can be achieved by carrying out an additional smoothing step. To this end, let $\ell(z_1, z_2)$ denote the kernel of a non-singular integral operator, L, on $L_2[0,1]^r$. That is, L is defined by

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

3. Non-parametric 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, Horowitz and Spokoiny (2001, 2002) and Guerre and Lavergne (2002).

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

$$S(x,z) \equiv T(g-G)(x,z) = 0, \tag{2.6}$$

for almost every $(x,z) \in [0,1]^{p+r}$. H_1 is equivalent to the statement that (2.6) does not hold on a set $B \subset [0,1]^{p+r}$ with non-zero Lebesgue measure. The test statistic is based on a sample analogue of $\int S(x,z)^2 dxdz$.

The motivation for basing a test of H_0 on $S(x,z)$ can be understood by observing that $g(x,z) = T_z^{-1}Q(x,z)$, where $Q(x,z) = f_Z(z)\boldsymbol{E}_{W|Z}[\boldsymbol{E}(Y \mid Z=z, W) f_{XZW}(x,z,W) \mid Z=z)]$, and f_Z is the probability density function of Z (Hall and Horowitz, 2005). T_z^{-1} is a discontinuous operator, and this discontinuity is the source of the ill-posed inverse problem in estimating g. Basing the test of H_0 on $S(x,z)$ avoids this problem because $S(x,z) = L(Q - T_zG)(x,z)$, which does not involve T_z^{-1}.

To form a sample analogue of $S(x,z)$, observe that $S(x,z) = \boldsymbol{E}\{[Y - G(X,Z)] \times f_{XW}(x, z, W)\ell(Z,z)\}$. Therefore, the analogue can be formed by replacing G and f_{XW} with estimates and $\boldsymbol{E}$ with the sample average in $\boldsymbol{E}\{[Y - G(X,Z)] f_{XW}(x,z,W)\ell(Z,z)\}$. To do this, let $\hat{f}^{(-i)}_{XZW}$ and $\hat{G}^{(-i)}$, respectively, denote leave-observation-i-out "boundary kernel" estimators of f_{XZW} and G (Gasser and Müller, 1979; Gasser, Müller and Mammitzsch, 1985). To describe these estimators, let $K_h(\cdot,\cdot)$ denote a boundary kernel function with the property that for all $\xi \in [0,1]$ and some integer $s \geq 2$

$$h^{-(j+1)} \int_{\xi}^{\xi+1} u^j K_h(u,\xi)du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq s-1. \end{cases} \tag{2.7}$$

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

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

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

Now define

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

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

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

and

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

where h_1 and h_2 are bandwidths, and

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

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

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

The test statistic is

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

H_0 is rejected if τ_n is large.

2.3. *Relation to testing for omitted variables*

As was mentioned in Section 1, an omitted variables test imposes stronger restrictions than are implied by the hypothesis of exogeneity. We now show that τ_n is not a test of whether W is an omitted variable in the mean regression of Y on (X,Z). Specifically, the null hypothesis of the τ_n test (that X is exogenous) can be true and the null hypothesis of the omitted variable test false simultaneously. The converse cannot occur. Therefore, the null hypothesis of the omitted variable test is more restrictive than the exogeneity hypothesis of the τ_n test.

In a test that W is an omitted variable, the null hypothesis is $\tilde{H}_0 : \boldsymbol{P}[\boldsymbol{E}(Y \mid X,Z,W) = \boldsymbol{E}(Y \mid X,Z)] = 1$. The alternative hypothesis is $\boldsymbol{P}[\boldsymbol{E}(Y \mid X,Z,W) = \boldsymbol{E}(Y \mid X,Z)] < 1$. Tests of $\tilde{H}_0$ have been developed by Gozalo (1993), Fan and Li (1996), Lavergne and Vuong (2000), and Aït-Sahalia, Bickel and Stoker (2001). The difference between these tests and τ_n is that τ_n assumes that $\boldsymbol{P}[\boldsymbol{E}(U \mid Z,W) = 0] = 1$ always and $\boldsymbol{P}[\boldsymbol{E}(U \mid X,Z) = 0] = 1$ if H_0 is true, but not that $\boldsymbol{P}[\boldsymbol{E}(U \mid X,Z,W) = 0] = 1$. It is easy to show that $\boldsymbol{E}(U \mid Z,W) = \boldsymbol{E}(U \mid X,Z) = 0$ with probability 1 does not imply that $\boldsymbol{E}(U \mid X,Z,W) = 0$ with probability 1. The exogeneity null allows the conditional mean of U given X and Z to vary with W. For example, let X, Z, W, and ν be independent random variables with means of 0, and set $U = XW + \nu$. Then $\boldsymbol{E}(U \mid Z,W) = \boldsymbol{E}(U \mid X,Z) = 0$ but $\boldsymbol{E}(U \mid X,Z,W) = XW$. The null hypothesis of the τ_n test is true, but the null hypothesis of the omitted variable test is false. Thus, τ_n is not a test of the hypothesis that W is an omitted variable. The hypothesis of exogeneity tested by τ_n is less restrictive than the hypothesis that W is an omitted variable.

3. ASYMPTOTIC PROPERTIES

3.1. *Regularity conditions*

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

1. (i) The support of (X, Z, W) is contained in $[0,1]^{2p+r}$. (ii) (X, Z, W) has a probability density function f_{XZW} with respect to Lebesgue measure. (iii) There is a constant $C_X > 0$ such that $f_{XZ}(x,z) \geq C_X$ for all $(x,z) \in \text{supp}(X, Z)$. (iv) There is a constant $C_f < \infty$ such that $|D_j f_{XZW}(x,z,w)| \leq C_f$ for all $(x,z,w) \in [0,1]^{2p+r}$ and $j = 0,1,\ldots,s$, where derivatives at the boundary of $\text{supp}(X, Z, W)$ are defined as one sided. (v) $|D_s f_{XZW}(x_1,z_1,w_1) - D_s f_{XZW}(x_2,z_2,w_2)| \leq C_f \|(x_1,z_1,w_1) - (x_2,z_2,w_2)\|$ for any s-th derivative and any (x_1,z_1,w_1), $(x_2,z_2,w_2) \in [0,1]^{2p+r}$. (vi) T_z is non-singular for almost every $z \in [0,1]^r$.
2. (i) $\boldsymbol{E}(U \mid Z = z, W = w) = 0$ and $\boldsymbol{E}(U^2 \mid Z = z, W = w) \leq C_{UV}$ for each $(z,w) \in [0,1]^{p+r}$, and some constant $C_{UV} < \infty$. (ii) $|g(x,z)| \leq C_g$ for some constant $C_g < \infty$ and all $(x,z) \in [0,1]^{p+r}$.
3. (i) The conditional mean function G satisfies $|D_j G(x,z)| \leq C_f$ for all $(x,z) \in [0,1]^{p+r}$ and $j = 0,1,\ldots,s$. (ii) $|D_s G(x_1,z_1) - D_s G(x_2,z_2)| \leq C_f \|(x_1,z_1) - (x_2,z_2)\|$ for any s-th derivative and any $(x_1,z_1,x_2,z_2) \in [0,1]^{2(p+r)}$. (iii) $\boldsymbol{E}(V^2 \mid X = x, Z = z) \leq C_{UV}$ for each $(x,z) \in [0,1]^{p+r}$.
4. (i) K_h satisfies (2.7) and $|K_h(u_2,\xi) - K_h(u_1,\xi)| \leq C_K |u_2 - u_1|/h$ for all u_2, u_1, all $\xi \in [0,1]$, and some constant $C_K < \infty$. For each $\xi \in [0,1]$, $K_h(h,\xi)$ is supported on $[(\xi-1)/h, \xi/h] \cap \mathcal{K}$, where $\mathcal{K}$ is a compact interval not depending on ξ. Moreover,

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

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

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

3.2. *Asymptotic properties of the test statistic*

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

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

4. T_z is a self-adjoint, positive-semi-definite operator, so its eigenvalues are non-negative. Under 1(vi), T_z is positive definite, and all its eigenvalues are strictly positive. If 1(vi) does not hold, then some eigenvalues are 0. Let A denote the linear space spanned by the eigenvectors of T_z corresponding to non-zero eigenvalues. If 1(vi) does not hold, then one can test for deviations from H_0 such that $g(\cdot,z) - G(\cdot,z)$ has a non-zero projection into A. It is not possible to test for deviations from H_0 for which $g(\cdot,z) - G(\cdot,z)$ lies entirely in the complement of A. In non-parametric IV estimation, validity of the instruments is equivalent to non-singularity of T_z. If T_z is non-singular, then whether the instruments are "weak" or "strong" depends on the rate at which the eigenvalues of T_z converge to 0. The instruments are weak if the eigenvalues converge rapidly and strong otherwise. There appears to be no simple, intuitive characterization of strength or weakness of instruments in this setting. In particular, in non-parametric IV estimation, the strength of correlation of X and W does not characterize the strength or weakness of W as an instrument.

and

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

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

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

Let $\{\omega_j : j = 1, 2, \ldots\}$ denote the eigenvalues of Ω sorted so that $\omega_1 \geq \omega_2 \geq \cdots \geq 0$.[5] Let $\{\chi^2_{1j} : j = 1, 2, \ldots\}$ 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.[6]

Theorem 1. *Let H_0 be true. Then under assumptions* 1–4,

$$\tau_n \to^d \sum_{j=1}^{\infty} \omega_j \chi^2_{1j}.$$

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

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

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

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

uniformly over t. Define

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

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

6. A referee asked whether τ_n satisfies the Liapounov condition (Serfling, 1980, p. 30), which would imply that τ_n is asymptotically normal. The answer is that τ_n does not satisfy the Liapounov condition. In our setting, the condition is $a_n/b_n \to 0$ as $n \to \infty$, where $a_n = \left(\sum_{j=1}^n \omega_j^\nu\right)^{1/\nu}$ and $b_n = \left(\sum_{j=1}^n \omega_j^2\right)^{1/2}$ for some $\nu > 2$. Boundedness of f_{XZW} implies that $0 < \sum_{j=1}^{\infty} \omega_j^c < \infty$ for any $c \geq 2$, so the convergence to 0 required by the Liapounov condition does not happen.

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 $\hat{\omega}_j \quad (j = 1, 2, \ldots, K_\varepsilon)$ be a consistent estimator of ω_j under H_0. Then the approximate critical value of τ_n is the $1-\alpha$ quantile of the distribution of

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

This quantile can be estimated with arbitrary accuracy by simulation.

At the cost of additional analytic complexity, it may be possible to let $\varepsilon \to 0$ and $K_\varepsilon \to \infty$ as $n \to \infty$, thereby obtaining a consistent estimator of the asymptotic critical value of τ_n. However, this would likely require stronger assumptions than are made here while providing little insight into the accuracy of the estimator or the choice of K_ε in applications. This is because the difference between the distributions of $\hat{\tau}_n$ and $\tilde{\tau}$ is a complicated function of the spacings and multiplicities of the ω_j's (Hall and Horowitz, 2006). The spacings and multiplicities are unknown in applications and appear difficult to estimate reliably.

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 $\{\hat{\omega}_j\}$. Because $V = U$ under H_0, a consistent estimator of $R(x_1, z_1; x_2, z_2)$ can be obtained by replacing unknown quantities with estimators on the R.H.S. of

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

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

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

Estimate the V_i's by

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

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

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

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

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

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

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

To obtain an accurate numerical approximation to the $\hat{\omega}_j$'s, let $\hat{F}(x,z)$ denote the $n \times 1$ vector whose i-th component is $[\hat{f}_{XZW}(x, Z_i, W_i) - \hat{t}_{Z_i}(X_i, x)/\hat{f}_{XW}(X_i, Z_i)]\ell(Z_i, z)$, and let Υ denote the $n \times n$ diagonal matrix whose (i,i) element is $\hat{V}_i^2$. Then

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

The computation of the eigenvalues can now be reduced to finding the eigenvalues of a finite-dimensional matrix. To this end, let $\{\phi_j : j = 1,2,\ldots\}$ be a complete, orthonormal basis for $L_2[0,1]^{p+r}$. Then

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

where

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

and

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

where

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

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

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

and

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

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

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

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

3.4. *Consistency of the test against a fixed alternative model*

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

Theorem 3. *Suppose that*

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

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

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

Because T is non-singular, 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 0.

3.5. *Asymptotic distribution under local alternatives*

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

$$Y = g(X, Z) + U; \boldsymbol{E}(U \mid Z, W) = 0; \quad \boldsymbol{E}(U \mid X, Z) = n^{-1/2} \Delta(X, Z), \tag{3.1}$$

where Δ is a bounded function on $[0,1]^{p+r}$. Under (3.1), the distributions of U and V depend on n, $n^{1/2}(U - V) = \Delta(X, Z)$, and $G(X, Z) = g(X, Z) + n^{-1/2} \Delta(X, Z)$. To provide a complete characterization of the sequence of alternative hypotheses, it is necessary to specify the dependence of the distributions of U and V on n. Here, it is assumed that

$$V = \nu + n^{-1/2} \varepsilon, \tag{3.2}$$

where ε and ν are random variables whose distributions do not depend on n, $\boldsymbol{E}(\nu \mid X, Z) = E(\nu \mid Z, W) = 0$, $Var(\nu) < \infty$, $\boldsymbol{E}(\varepsilon \mid X, Z) = 0$, $\boldsymbol{E}(\varepsilon \mid Z, W) = -\boldsymbol{E}[\Delta(X, Z) \mid Z, W]$, and $\mathrm{Var}(\varepsilon) < \infty$. It follows from (3.1) and (3.2) that

$$U = \nu + n^{-1/2} \Delta(X, Z) + n^{-1/2} \varepsilon. \tag{3.3}$$

The following additional notation is used. Define

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

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

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

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

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

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

Theorem 4. *Let assumptions* 1–4 *hold. Under the sequence of local alternatives* (3.1)–(3.3),

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

It follows from Theorems 2 and 4 that under (3.1)–(3.3),

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

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

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

The following theorem states the result of this section.

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

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

and

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

3.7. *Alternative weights*

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

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

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

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

where

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

If $H(z,w) = \tilde{H}(zw)$ for a suitably chosen function $\tilde{H}$, then τ_{nH} is a modification of the statistic of Bierens (1990) and Bierens and Ploberger (1997) for testing the hypothesis that a conditional mean function belongs to a specified, finite-dimensional parametric family. In this section, it is shown that the power of the τ_{nH} test can be low relative to that of the τ_n test. Specifically, there are combinations of density functions f_{XW} and local alternative models (3.1)–(3.3) such that an α-level τ_{nH} test based on a fixed H that does not depend on the sampled population has asymptotic local power arbitrarily close to α, whereas the α-level τ_n test has asymptotic local power that is bounded away from α. The opposite situation cannot occur under the assumptions of this paper. That is, it is not possible for the asymptotic power of the α-level τ_n test to approach α while the power of the α-level τ_{nH} test remains bounded away from α.

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

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

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

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

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

and

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

Let $\{\bar{\omega}_j, \bar{\psi}_j : j = 1, 2, \ldots\}$ and $\{\bar{\omega}_{jH}, \bar{\psi}_{jH} : j = 1, 2, \ldots\}$ denote the eigenvalues and eigenvectors of $\bar{\Omega}$ and $\bar{\Omega}_H$, respectively, with the eigenvalues sorted in decreasing order. For Δ defined as in (3.1), define $\bar{\mu}(z) = (T\Delta)(z)$,

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

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

and

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

Then arguments like those used to prove Theorem 4 show that under the sequence of local alternatives (3.1)–(3.3) with a known function G,

$$\tau_n \to^d \sum_{j=1}^{\infty} \bar{\omega}_j \chi^2_{1j}(\bar{\mu}_j^2/\bar{\omega}_j),$$

and

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

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

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

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

Let

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

Let $\boldsymbol{E}(U^2 \mid W = w) = 1$ for all $w \in [0,1]$. Then $\bar{R}(z_1, z_2) = t(z_1, z_2)$, $\bar{\omega}_j = \lambda_j$, and $\sum_{j=1}^{\infty} \bar{\omega}_j$ is non-zero and finite. Set $\Delta(x) = D\phi_\ell(x)$ for some finite $D > 0$. Then $\|\bar{\mu}\|^2 = D^2\lambda_\ell^2 = D^2$. Since H is fixed, it suffices to show that ℓ can be chosen so that $\|\bar{\mu}_H\|^2$ is arbitrarily close to 0. To do this, observe that $H(z,w)$ has the Fourier representation

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

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

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

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

It follows from the Cauchy–Schwartz inequality that

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

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

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) were generated by $X = \Phi(\xi)$ and $W = \Phi(\zeta)$, where Φ is the cumulative normal distribution function, $\zeta \sim N(0,1)$, $\xi = \rho\zeta + (1-\rho^2)^{1/2}\varepsilon$, $\varepsilon \sim N(0,1)$, and $\rho = 0{\cdot}35$ or $\rho = 0{\cdot}7$, depending on the experiment. Realizations of Y were generated from

$$Y = \theta_0 + \theta_1 X + \sigma_U U, \tag{4.1}$$

TABLE 1

Results of Monte Carlo experiments

n	η	τ_n	Empirical probability that H_0 is rejected using Hausman test	τ_{nH}
$\rho = 0{\cdot}35$				
250	0·0	0·042	0·050	0·012
	0·10	0·062	0·072	0·022
	0·15	0·077	0·126	0·039
	0·20	0·076	0·164	0·068
	0·25	0·119	0·265	0·116
500	0·0	0·048	0·055	0·025
	0·10	0·256	0·304	0·187
	0·15	0·539	0·590	0·429
	0·20	0·814	0·876	0·724
	0·25	0·945	0·971	0·922
750	0·0	0·048	0·053	0·035
	0·10	0·137	0·172	0·131
	0·15	0·274	0·313	0·232
	0·20	0·422	0·468	0·379
	0·25	0·596	0·675	0·601
$\rho = 0{\cdot}70$				
250	0·0	0·047	0·051	0·028
	0·10	0·156	0·188	0·079
	0·15	0·293	0·366	0·192
	0·20	0·464	0·568	0·360
	0·25	0·705	0·802	0·563
500	0·0	0·048	0·055	0·025
	0·10	0·256	0·304	0·187
	0·15	0·539	0·590	0·429
	0·20	0·814	0·876	0·724
	0·25	0·945	0·971	0·922
750	0·0	0·050	0·049	0·025
	0·10	0·383	0·479	0·298
	0·15	0·728	0·806	0·646
	0·20	0·929	0·958	0·896
	0·25	0·994	0·997	0·983

where $\theta_0 = 0$, $\theta_1 = 0{\cdot}5$, $U = \eta\varepsilon + (1-\eta^2)^{1/2}\nu$, $\nu \sim N(0,1)$, $\sigma_U = 0{\cdot}2$, and η is a constant parameter whose value varies among experiments. H_0 is true if $\eta = 0$ and false otherwise. To provide a basis for judging whether the power of the τ_n test is high or low, we also report the results of a Hausman (1978) type test of the hypothesis that the OLS and IV estimators of θ_1 in (4.1) are equal. The instruments used for IV estimation of (4.1) are $(1, W)$. In addition, we report the results of simulations with τ_{nH}. The weight function is $H(x,w) = \exp(xw)$ and is taken from Bierens (1990). The bandwidth used to estimate f_{XW} was selected by cross-validation. The bandwidth used to estimate f_X is $n^{1/5-7/24}$ times the cross-validation bandwidth. The kernel is $K(v) = (15/16)(1-v^2)^2 I(|v| \le 1)$, where I is the indicator function. The asymptotic critical value was estimated by setting $K_\varepsilon = 25$. The results of the experiments are not sensitive to the choice of K_ε, and the estimated eigenvalues $\hat{\omega}_j$ are very close to 0 when $j > 25$. The experiments use sample sizes of n =250, 500, and 750 and the nominal 0·05 level. There are 1000 Monte Carlo replications in each experiment.

The results of the experiments are shown in Table 1. The differences between the nominal and empirical rejection probabilities of the τ_n and Hausman-type tests are small when H_0 is true. When H_0 is false, the power of the τ_n test is, not surprisingly, somewhat smaller than the power of the Hausman-type test, which is parametric, but the differences in power are not great. The performance of τ_{nH} is worse than that of τ_n. When H_0 is true, the difference between the nominal and empirical rejection probabilities of the τ_{nH} test is relatively large, and the power of the τ_{nH} test is usually lower than that of the τ_n test.

5. EXOGENEITY AND CONSUMER EXPANSION PATHS

The empirical analysis of consumer expansion paths (or Engel curves) concerns the relationship between expenditures on specific commodities and total consumption; see Deaton (1998), for example. The shape of the expansion path defines whether a good is a necessity or a luxury, and knowledge of the expansion path allows the researcher to measure reactions to policies that change the resources allocated to individuals across the wealth distribution. The relationship we wish to recover for policy analysis is the structural function that describes changes in commodity demands in response to exogenous changes in overall consumption. In household expenditure survey data it is likely that the total consumption (expenditure) variable will be endogenous for the expansion path, and standard non-parametric regression estimators will not recover the structural relationship of interest.

Here we explore the expansion path relationship for leisure services bought by consumers in a particular month. This curve has been shown to be non-linear in non-parametric regression analysis (see Blundell, Duncan and Pendakur, 1998) and so poses a particularly acute problem in estimation under endogeneity. We present an empirical application of our test statistic τ_n to this expansion path problem and assess whether we can reject the exogeneity hypothesis for total consumption in the leisure services expansion path. The curve is given by (2.3) with $p = 1$ and $r = 0$, where Y denotes the expenditure share of services, X denotes the logarithm of total expenditures, and W denotes annual income from wages and salaries of the head of household.

The data consist of household-level observations from the British Family Expenditure Survey, which is a popular data source for studying consumer behaviour.[7] This is a diary-based household survey that is supplemented by recall information. We use a subsample of 1518 married couples with one or two children and an employed head of household.[8] W should be a good instrument for X if income from wages and salaries is not influenced by household budgeting decisions.

The bandwidths for estimating f_{XW} were selected by the method described in the Monte Carlo section. The kernel is the same as the one used in the Monte Carlo experiments. As in the experiments, the critical value of τ_n was estimated by setting $K_\varepsilon = 25$. The τ_n test of the hypothesis that X is exogenous gives $\tau_n = 0{\cdot}162$ with a 0·05-level critical value of 0·151. Thus, the test rejects the hypothesis that X is exogenous.

Parametric specifications are often linear or quadratic in X (Muellbauer, 1976; Banks, Blundell and Lewbel, 1997). Consequently, the hypothesis was also tested by comparing the OLS and IV estimates of θ_1 and θ_2 in the quadratic model

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

The instruments are $(1, W, W^2)$. The hypothesis that the OLS estimates of θ_1 and θ_2 equal the IV estimates is rejected at the 0·05 level. Thus, the τ_n test and the parametric test both reject the hypothesis that the logarithm of total expenditures is exogenous.

7. See Blundell, Pashardes and Weber (1993), for example.

8. The data are available on the *Review*'s website. This is also the sample selection used in the Blundell *et al.* (2007) study.

6. CONCLUSIONS

Endogeneity of explanatory variables is an important problem in applied econometrics. Erroneously assuming that explanatory variables are exogenous can cause estimation results to be highly misleading. Conversely, unnecessarily assuming that one or more variables are endogenous can greatly reduce estimation precision, especially in the non-parametric setting considered in this paper. This paper has described a test for exogeneity of explanatory variables that minimizes the need for auxiliary assumptions that are not required by the definition of exogeneity. Specifically, the test does not make parametric functional form assumptions, thereby avoiding the possibility of obtaining a misleading result due to model misspecification. In addition, the test described here does not make the auxiliary assumptions that are implied by a test for omitted IV. We have shown that the hypothesis of exogeneity can be true and the hypothesis of no omitted variables false simultaneously. The opposite situation cannot occur. Thus, an omitted variables test requires assumptions that are stronger than implied by exogeneity and can erroneously reject the hypothesis of exogeneity due to failure of the auxiliary conditions to hold. We have illustrated the usefulness of the new exogeneity test through Monte Carlo experiments and an application to estimation of Engel curves.

APPENDIX : PROOFS OF THEOREMS

To minimize the complexity of the presentation, we assume that $p = 1$, $r = 0$, and $s = 2$. The proofs for $p > 1$, $r > 0$, and/or $s > 2$ are identical after replacing quantities for $p = 1$, $r = 0$, and $s = 2$ with the analogous quantities for the more general case. Let f_{XW} denote the density function of (X, W).

Define

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

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

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

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

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

and

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

Then

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

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

Lemma A1 *As $n \to \infty$,*

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

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

Proof. Define

$$R_{n1}^{(-i)}(x) = \frac{1}{nh_2 f_X(x)} \sum_{j=1, j\neq i}^{n} V_j K_{h_2}(x - X_j, x),$$

$$R_{n2}^{(-i)}(x) = \frac{1}{nh_2 f_X(x)} \sum_{j=1, j\neq i}^{n} [G(X_j) - G(x)] K_{h_2}(x - X_j, x),$$

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

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

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

and

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

Standard calculations for kernel estimators show that

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

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

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

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

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

as $n \to \infty$. Therefore,

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

and

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

by Markov's inequality. Moreover, we can write

$$E_i [R_{n1}^{(-i)}(X_i) f_{XW}(z, W_i)] = \frac{1}{nh_2} \sum_{j=1, j\neq i}^{n} V_j \int_0^1 [f_{XW}(x, w) f_{XW}(z, w)/f_X(x)] K_{h_2}(x - X_j, x) dx dw$$

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

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

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

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

Lemma A2 *As $n \to \infty$, $\int_0^1 S_{n4}^2(z)dz = o_p(1)$.*

Proof. Define

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

Then

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

$$= D_n + o(1). \tag{A.4}$$

Now define

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

and

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

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

$$D_{n1} = n^{-1} E \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} U_i U_j \int_0^1 [\hat{f}_{XW}^{(-i,-j)}(z, W_i) - f_{XW}(z, W_i)][\hat{f}_{XW}^{(-j,-i)}(z, W_i) - f_{XW}(z, W_j)]dz$$

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

and

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

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

Lemma A3 *As $n \to \infty$, $S_{n6}(z) = o_p(1)$ uniformly over $z \in [0, 1]$.*

Proof. This follows from $\hat{f}_{XW}^{(-i)}(x, w) - f_{XW}(x, w) = O[(\log n)/(nh_1^2)^{1/2} + h_1^2]$ a.s. uniformly over $(x, w) \in [0, 1]^2$ and $\hat{G}^{(-i)}(x) - G(x) = O[(\log n)/(nh_2)^{1/2} + h_2^2]$ a.s. uniformly over $x \in [0, 1]$. ‖

Proof of Theorem 1. Under H_0, $S_{n2}(z) = S_{n5}(z) = 0$ for all $z \in [0, 1]$. Therefore, it follows from Lemmas A1–A3 that

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

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

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

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

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

This will be done by proving that

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

To do this, observe that by a uniform law of large numbers of Pakes and Pollard (1989, lemma 2.8), $n^{-1/2}S_{n2}(z) = (Tq)(z) + o_p(1)$ uniformly over $z \in [0,1]$. Moreover, $n^{-1/2}S_{n5}(z) = o_p(1)$ uniformly over $z \in [0,1]$ because $\hat{f}_{XW}^{(-i)}(z,w) - f_{XW}(z,w) = O[(\log n)/(nh_1^2)^{1/2} + h_1^2]$ a.s. uniformly over $(z,w) \in [0,1]^2$. Combining these results with Lemmas A1–A3 yields

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

where $\int_0^1 r_n^2(z)dz = o_p(1)$ as $n \to \infty$. It follows from Theorem 1 that $n^{-1}\int_0^1 B_n^2(z)dz = o_p(1)$. Therefore, $n^{-1}\tau_n \to^p \int_0^1 [(Tq)(z)]^2 dz$. ‖

Proof of Theorem 4. The conclusions of Lemmas A1–A3 hold under (3.1)–(3.3). Therefore,

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

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

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

a.s. uniformly over z. In addition

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

a.s. uniformly over z. Therefore, $S_n(z) = B_n(z) + \mu(z) + r_n(z)$. But

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

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

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

where

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

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

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

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

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

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

Proof of Theorem 5. Let z_α denote the critical value of τ_n. Observe that z_α is bounded uniformly over $\mathcal{F}_{nc}$. The arguments used to prove Lemmas A1–A3 show that $\int_0^1 S_{nj}^2(z)dz = o_p(1)$ for $j = 4, 6$ and $\int_0^1 S_{n3}^2(z)dz = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. In addition, an application of Markov's inequality shows that $\int_0^1 S_{n1}^2(z)dz = O_p(1)$ uniformly over $\mathcal{F}_{nc}$. Define

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

and

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

Let $\|\cdot\|$ denote the $L_2[0,1]$ norm. Use the inequality $a^2 \geq 0{\cdot}5b^2 - (b-a)^2$ with $a = S_n$ and $b = S_{n2} + S_{n5}$ to obtain

$$P(\tau_n > z_\alpha) \geq P(0{\cdot}5\|D_n\|^2 - \|\tilde{S}_n\|^2 > z_\alpha).$$

For any finite $M > 0$,

$$\begin{aligned} P(0{\cdot}5\|D_n\|^2 - \|\tilde{S}_n\|^2 \leq z_\alpha) &= P(0{\cdot}5\|D_n\|^2 \leq z_\alpha + \|\tilde{S}_n\|^2, \|\tilde{S}_n\|^2 \leq M) \\ &\quad + P(0{\cdot}5\|D_n\|^2 \leq z_\alpha + \|\tilde{S}_n\|^2, \|\tilde{S}_n\|^2 > M) \\ &\leq P(0{\cdot}5\|D_n\|^2 \leq z_\alpha + M) + P(\|\tilde{S}_n\|^2 > M). \end{aligned}$$

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

$$P(0{\cdot}5\|D_n\|^2 - \|\tilde{S}_n\|^2 \leq z_\alpha) \leq P(0{\cdot}5\|D_n\|^2 \leq z_\alpha + M) + \varepsilon,$$

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

$$P(0{\cdot}5\|D_n\|^2 - \|\tilde{S}_n\|^2 > z_\alpha) \geq P(0{\cdot}5\|D_n\|^2 > z_\alpha + M) - \varepsilon,$$

and

$$P(\tau_n > z_\alpha) \geq P(0{\cdot}5\|D_n\|^2 > z_\alpha + M) - \varepsilon. \tag{A.5}$$

Now

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

Therefore,

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

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

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

and

$$E D_n(z) \geq 0{\cdot}5n^{1/2}(Tq)(z),$$

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

$$\begin{aligned} D_n(z) - E D_n(z) &= n^{-1/2} \sum_{i=1}^n [q(X_i) E^{(-i)} \hat{f}_{XW}^{(-i)}(z, W_i) - Eq(X) \hat{f}_{XW}^{(-i)}(z, W_i)] \\ &\quad + n^{-1/2} \sum_{i=1}^n q(X_i)[f_{XW}^{(-i)}(z, W_i) - E^{(-i)} \hat{f}_{XW}^{(-i)}(z, W_i)] \\ &\equiv D_{n1}(z) + D_{n2}(z), \end{aligned}$$

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

$$|D_{n2}(z)| \leq \frac{r_n \log n}{nh_1} \sum_{i=1}^{n} |q(X_i)| = \frac{r_n \log n}{h_1} [E|q(X)| + O_p(n^{-1/2})],$$

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

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

A further application of $a^2 \geq 0{\cdot}5b^2 - (b-a)^2$ with $a = D_n$ and $b = ED_n$ gives

$$\|D_n\|^2 \geq n\|Tq\|^2 \left[0{\cdot}125 - \left(\frac{r_n \log n}{n^{1/2} h_1^3}\right)^2 \frac{h_1^4 (E|q|)^2}{\|Tq\|^2} O_p(1)\right] + O_p(1)$$

$$= n\|Tq\|^2 \left[0{\cdot}125 - \left(\frac{r_n \log n}{n^{1/2} h_1^3}\right)^2 o_p(1)\right] + O_p(1)$$

uniformly over $\mathcal{F}_{nc}$. Therefore, if C is sufficiently large, $0{\cdot}5\|D_n\|^2 > z_\alpha + M$ with probability approaching 1 as $n \to \infty$ uniformly over $\mathcal{F}_{nc}$. ‖

Acknowledgements. We thank Xiaohong Chen, James J. Heckman, the editor, and two anonymous referees for comments. Part of this research was carried out while Joel Horowitz was a visitor at the Centre for Microdata Methods and Practice, University College London and Institute for Fiscal Studies. Horowitz's research was also supported in part by NSF Grant SES 0352675. Richard Blundell thanks the ESRC Centre for the Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies for financial support.

REFERENCES

AïT-SAHALIA, Y., BICKEL, P. J. and STOKER, T. M. (2001), "Goodness of Fit Tests for Kernel Regression with Application to Option Implied Volatilities", *Journal of Econometrics*, **105**, 363–412.

BANKS, J., BLUNDELL, R. W. and LEWBEL, A. (1997), "Quadratic Engel Curves, Indirect Tax Reform and Welfare Measurement", *Review of Economics and Statistics*, **79**, 527–539.

BHATIA, R., DAVIS, C. and MCINTOSH, A. (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 PLOBERGER, W. (1997), "Asymptotic Theory of Integrated Conditional Moment Tests", *Econometrica*, **65**, 1129–1151.

BLUNDELL, R., BROWNING, M. and CRAWFORD, I. A. (2003), "Nonparametric Engel Curves and Revealed Preference", *Econometrica*, **71**, 205–240.

BLUNDELL, R., CHEN, X. and KRISTENSEN, D. (2007), "Semi-Nonparametric IV Estimation of Shape Invariant Engle Curves" *Econometrica* (forthcoming).

BLUNDELL, R., DUNCAN, A. and PENDAKUR, K. (1998), "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, **13**, 435–461.

BLUNDELL, R., PASHARDES, P. and WEBER, G. (1993), "What Do We Learn About Consumer Demand Patterns from Micro Data?", *American Economic Review*, **83**, 570–597.

BLUNDELL, R. and POWELL, J. (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models"', in M. Dewatripont, L. Hansen and S. J. Turnsovsky (eds.) *Advances in Economics and Econometrics*, Ch. 8 (Cambridge, UK: Cambridge University Press) ESM 36, 312–357.

CARRASCO, M., FLORENS, J.-P. and RENAULT, E. (2005), "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization", in E. E. Leamer and J. J. Heckman (eds.) *Handbook of Econometrics*, Vol. 6 (Amsterdam: North-Holland).

DAROLLES, S., FLORENS, J.-P. and RENAULT, E. (2006), "Nonparametric Instrumental Regression" (Working Paper, GREMAQ, University of Social Science, Toulouse).

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

FAN, Y. and LI, Q. (1996), "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms", *Econometrica*, **64**, 865–890.

GASSER, T. and MÜLLER, H. G. (1979), "Kernel Estimation of Regression Functions", in T. Gasser and M. Rosenblatt (eds.) *Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics*, Vol. 757 (New York: Springer) 23–68.

GASSER, T., MÜLLER, H. G. and MAMMITZSCH, V. (1985), "Kernels and Nonparametric Curve Estimation", *Journal of the Royal Statistical Society Series B*, **47**, 238–252.

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

GUERRE, E. and LAVERGNE, P. (2002), "Optimal Minimax Rates for Nonparametric Specification Testing in Regression Models", *Econometric Theory*, **18**, 1139–1171.

HALL, P. and HOROWITZ, J. L. (2005), "Nonparametric Methods for Inference in the Presence of Instrumental Variables", *Annals of Statistics*, **33**, 2904–2929.

HALL, P. and HOROWITZ, J. L. (2006), "Methodology and Convergence Rates for Functional Linear Regression", *Annals of Statistics* (forthcoming).

HAUSMAN, J. A. (1978), "Specification Tests in Econometrics", *Econometrica*, **46**, 1251–1271.

HOROWITZ, J. L. (2006), "Testing a Parametric Model Against a Nonparametric Alternative with Identification through Instrumental Variables", *Econometrica*, **74**, 521–528.

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

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

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

LAVERGNE, P. and VUONG, Q. (2000), "Nonparametric Significance Testing", *Econometric Theory*, **16**, 576–601.

MEULLBAUER, J. (1976), "Community Preferences and the Representative Consumer", *Econometrica*, **44**, 525–543.

NEWEY, W. K. and POWELL, J. L. (2003), "Instrumental Variable Estimation of Nonparametric Models", *Econometrica*, **71**, 1565–1578.

NEWEY, W. K., POWELL, J. L. and VELLA, F. (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, **67**, 565–603.

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

PAKES, A. and POLLARD, D. (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, **57**, 1027–1057.

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

SMITH, R. J. (1994), "Asymptotically Optimal Tests using Limited Information and Testing for Exogeneity", *Econometric Theory*, **10**, 53–69.