On the testability of identification in some nonparametric models with endogeneity

Ivan Canay
Andres Santos
Azeem Shaikh

On the Testability of Identification in Some Nonparametric Models with Endogeneity

Ivan A. Canay

Department of Economics

Northwestern University

iacanay@northwestern.edu

Andres Santos

Department of Economics

University of California - San Diego

a2santos@ucsd.edu

Azeem M. Shaikh*

Department of Economics

University of Chicago

amshaikh@uchicago.edu

June 7, 2012

Abstract

This paper examines three distinct hypothesis testing problems that arise in the context of identification of some nonparametric models with endogeneity. The first hypothesis testing problem we study concerns testing necessary conditions for identification in some nonparametric models with endogeneity involving mean independence restrictions. These conditions are typically referred to as completeness conditions. The second and third hypothesis testing problems we examine concern testing for identification directly in some nonparametric models with endogeneity involving quantile independence restrictions. For each of these hypothesis testing problems, we provide conditions under which *any* test will have power no greater than size against *any* alternative. In this sense, we conclude that no nontrivial tests for these hypothesis testing problems exist.

KEYWORDS: Instrumental Variables, Identification, Completeness, Bounded Completeness

1 Introduction

Instrumental variables (IV) methods have a prominent role in econometrics due to their ability to uncover causal effects in observational studies. Though traditionally parametric in nature, an important literature has extended IV methods to a variety of nonparametric settings. Among these extensions, of particular prominence is the additively separable specification in which for an outcome of interest Y, a regressor W, and an instrument Z it is assumed that

$$Y = \theta(W) + \epsilon , \tag{1}$$

with ϵ mean independent of Z. Under the maintained assumption that the model is correct, Newey and Powell (2003) showed identification of θ to be equivalent to the conditional distribution of W given Z satisfying a completeness condition. Complementing their prevalent use in statistics (Lehmann and Scheffé, 1950, 1955), completeness conditions have since then been widely used in econometrics – see Hall and Horowitz (2005), Blundell et al. (2007), Hu and Schennach (2008), Berry and Haile (2010a), Darolles et al. (2011) and the references therein.

Despite the theoretical importance of completeness conditions, little evidence has been provided for or against these assumptions being satisfied in datasets of interest to economists. We note, however, that since completeness conditions impose restrictions on the distribution of the observed data, it is potentially possible to provide such evidence by testing the validity of these assumptions. This paper explores precisely this possibility. Specifically, we study whether it is possible to test the null hypothesis that a completeness condition does not hold against the alternative that it does hold. Such a hypothesis testing problem is consistent with a setting in which a researcher wishes to assert the model is identified and hopes to find evidence in favor of this claim in the data. This setup is also analogous to tests of rank conditions in linear models with endogeneity, where the null hypothesis is that of rank-deficiency – see Remarks 3.1 and 3.4.

In this paper we show that, under commonly imposed restrictions on the distribution of the data, the null hypothesis that the completeness condition does not hold is in fact untestable. Formally, we establish that *any* test will have power no greater than size against *any* alternative. It is therefore not possible to provide empirical evidence in favor of the completeness condition by means of such a test. This conclusion is in contrast to the testability of a failure of the rank condition in linear specifications of θ, for which nontrivial tests do exist. Thus, while completeness conditions provide an intuitive generalization of the rank condition in a linear specification, we note that the empirical implications of these assumptions are substantially different.

We additionally derive analogous results in two other prevalent nonparametric models with endogeneity. The first such model follows the specification in (1) with a pre-specified conditional quantile of ϵ assumed independent of Z – see Chernozhukov and Hansen (2005), Horowitz and Lee (2007), Chernozhukov et al. (2010), and Chen and Pouzo (2012). The second such model

follows a specification in which θ is allowed to depend nonseparably on both W and ϵ, with the dependence on ϵ being monotonic, and all conditional quantiles of ϵ assumed independent of Z – see Chernozhukov and Hansen (2005), Imbens and Newey (2009), Torgovitsky (2011), and Berry and Haile (2009, 2010b). Due to the nonlinear nature of such models, simple, global rank conditions such as completeness conditions are unavailable. For this reason, we instead directly consider the testability of the null hypothesis that identification fails against the alternative hypothesis that it holds. In accord with our results concerning the testability of completeness conditions, we obtain conditions under which no nontrivial tests exist for these hypothesis testing problems either.

This paper contributes to an important literature on impossibility results in econometrics – see Leeb and Pötscher (2008) and Müller (2008) for recent examples, and Dufour (2003) for an excellent overview. First among such results is Bahadur and Savage (1956) who documented the impossibility of conducting nontrivial inference on the mean without appropriate restrictions on the data generating processes. Romano (2004) later showed that the Bahadur and Savage (1956) result is the consequence of the set of distributions satisfying the null hypothesis being dense in the set of distributions satisfying the alternative. Intuitively, if for every distribution in the alternative there exists an arbitrarily close distribution satisfying the null, then it will be impossible to discriminate between them from data – see Pötscher (2002) for related ideas in estimation. Our results share this intuition, but require novel arguments for showing that distributions for which a completeness condition or identification fails can approximate distributions for which it holds arbitrarily well.

We emphasize that our results should not be interpreted as an indictment against nonparametric methods in models with endogeneity. The nontestability of completeness conditions or identification does not imply these assumptions to be false. The impossibility of providing supporting empirical evidence by means of such a test, however, does suggest that it is prudent to justify their use with alternative arguments in favor of their validity. In situations where such arguments are not available, empirical researchers may employ statistical methods that do not rely on these assumptions. Recent work towards this end includes Chen et al. (2011), Freyberger and Horowitz (2012) and Santos (2012), who propose inferential procedures that allow for partial identification in these settings. Additionally, our results do not preclude the testability of completeness conditions or identification under alternative restrictions on θ or the distribution of the observed data. For instance, our arguments may not extend easily to settings where θ is restricted to be a density (Hoderlein et al., 2012) or is semiparametrically specified (Ai and Chen, 2003). We further discuss the possible implications of alternative restrictions on the data generating process in the text – see Remarks 2.1 and 3.3.

The remainder of the paper is organized as follows. Section 2 elaborates on the nature of our impossibility results, and reviews a general framework for deriving them. The main results are developed in Section 3, while Section 4 briefly concludes. All proofs are contained in the Appendix.

2 Setup

Before formally stating the three distinct hypothesis problems we consider, it will be useful to introduce some notation and elaborate on the nature of our impossibility results. Toward this end, we let $\{V_i\}_{i=1}^n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ and P^n denote the n-fold product $\bigotimes_{i=1}^n P$. The hypothesis testing problems we study may then be expressed as

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P}_1 \;, \tag{2}$$

where $\mathbf{P}_0$ is the subset of $\mathbf{P}$ for which the null hypothesis holds and $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$. For a sequence of (possibly randomized) tests $\{\phi_n\}_{n=1}^\infty$, the corresponding size at sample size n is then

$$\sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \;. \tag{3}$$

In our analysis, we will show that under commonly imposed restrictions on the set of distributions $\mathbf{P}$, the three hypothesis testing problems we examine share the property that

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \tag{4}$$

for any sequence of (possibly randomized) tests $\{\phi_n\}_{n=1}^\infty$ and any sample size n. Equivalently, result (4) establishes that for *all* tests, the power against *any* alternative $P \in \mathbf{P}_1$ is always bounded above by the size of the test. It also follows from such an assertion, that any sequence of (possibly randomized) tests $\{\phi_n\}_{n=1}^\infty$ that controls asymptotic size at level $\alpha \in (0, 1)$ will have asymptotic power no larger than α against any alternative. Formally, (4) immediately yields that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \leq \alpha \implies \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \alpha \;. \tag{5}$$

We therefore conclude that no nontrivial test exists for hypothesis testing problems satisfying property (4). In other words, in such settings no test can outperform a procedure that simply ignores the data and randomly rejects with a prespecified probability.

2.1 A Useful Lemma

Underlying our arguments is a powerful result originally found in Romano (2004), which we restate due to its importance in our derivations. In the statement of the lemma, $\|P - Q\|_{TV}$ denotes the Total Variation distance between probability measures P and Q – e.g., for $\lambda \equiv (P + Q)/2$

$$\|P - Q\|_{TV} \equiv \frac{1}{2} \int \left| \frac{dQ}{d\lambda} - \frac{dP}{d\lambda} \right| d\lambda \;. \tag{6}$$

4

Lemma 2.1. *Let* $\mathbf{M}$ *denote the space of Borel probability measures on a separable metric space* $\mathbf{V}$. *Suppose* $\mathbf{P} \subseteq \mathbf{M}$ *and* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *satisfy* $\mathbf{P} = \mathbf{P}_0 \cup \mathbf{P}_1$. *If for each* $P \in \mathbf{P}_1$ *there exists a sequence* $\{P_k\}_{k=1}^{\infty}$ *in* $\mathbf{P}_0$ *with* $\|P - P_k\|_{TV} = o(1)$, *then every sequence of test functions* $\{\phi_n\}_{n=1}^{\infty}$ *satisfies*

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \ \ \textit{for all } n \geq 1 \ . \tag{7}$$

Heuristically, Lemma 2.1 states that if each $P \in \mathbf{P}_1$ is on the boundary of the set of distributions satisfying the null hypothesis, then, by continuity, the probability of rejection under any $P \in \mathbf{P}_1$ must be no larger than the size of the test. Theorem 1 in Romano (2004) establishes that the appropriate topology for this purpose is that induced by the Total Variation distance. A metric compatible with weak convergence, such as the Lévy-Prokhorov metric, would be insufficient as it would not guarantee convergence of integrals defining rejection probabilities. By contrast, the Total Variation distance between two measures P and Q is intimately related to the statistical properties of the best test for distinguishing between P and Q, (LeCam, 1986). For this reason, some authors have referred to the Total Variation distance as the testing metric (Donoho, 1988).

In each of the three hypothesis testing problems we consider, we establish nonexistence of nontrivial tests by constructing for each $P \in \mathbf{P}_1$ a sequence $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ with $\|P - P_k\|_{TV} = o(1)$ and applying Lemma 2.1. In this way, our results are driven by $\mathbf{P}_0$ being dense in $\mathbf{P}_1$ with respect to $\|\cdot\|_{TV}$ in all three settings we examine. It is worth emphasizing that the Total Variation metric plays no role in determining $\mathbf{P}$ nor the null and alternative hypothesis. Indeed, $\mathbf{P}$ may posses other natural topologies, such as that induced by the Euclidean metric in parametric models, or that induced by a Hölder norm in sets of smooth densities. However, Lemma 2.1 reveals that, regardless of what topology $\mathbf{P}$ originally has, for our purposes we must examine $\mathbf{P}$ under the topology induced by the Total Variation metric – see also Remarks 2.1 and 3.3.

Remark 2.1. The converse of Lemma 2.1 implies that in order for nontrivial tests to exist, there must be a $P \in \mathbf{P}_1$ that is not in the closure of $\mathbf{P}_0$ with respect to $\|\cdot\|_{TV}$. For example, for some functional $\Upsilon : \mathbf{P} \to \mathbf{R}$ that is continuous with respect to a metric $\|\cdot\|_{\mathbf{P}}$ consider testing (2) with

$$\mathbf{P}_0 = \{P \in \mathbf{P} : \Upsilon(P) = 0\} \ , \tag{8}$$

and $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$. If $\mathbf{P}$ is restricted to be a metric space under $\|\cdot\|_{TV}$ and compact under $\|\cdot\|_{\mathbf{P}}$, then no $P \in \mathbf{P}_1$ is in the closure of $\mathbf{P}_0$ with respect to $\|\cdot\|_{TV}$. For instance, when conducting inference on the level of a density at a point, the Arzlá-Ascoli theorem implies it suffices to identify $\mathbf{P}$ with a family of bounded equicontinuous densities with uniformly bounded support. Unfortunately, finding appropriate restrictions on $\mathbf{P}$ for the settings we examine is far more challenging. In particular, smoothness restrictions on densities are not sufficient for this purpose – see Remark 3.3. ∎

3   Main Results

In this section, we show for each of the three hypothesis testing problems we consider that the power of any test against any alternative is always bounded above by the size of the test, i.e., we show (4) holds. As mentioned previously, these results further imply that any test that controls asymptotic size will have trivial asymptotic power against any alternative, i.e., they imply (5).

3.1   Testing Completeness

In order to formally state the first hypothesis testing problem we consider, let $V_i = (X_i, Z_i) \in \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ be random variables distributed according to $P \in \mathbf{P}$. For $Z_i = (Z_i^{(1)}, Z_i^{(2)})$, with the subvector $Z_i^{(1)}$ possibly empty, we let $W_i = (X_i, Z_i^{(1)}) \in \mathbf{R}^{d_w}$, and consider the condition

$$E_P[\theta(W_i)|Z_i] = 0 \ P\text{-a.s. for } \theta \in \Theta(P) \implies \theta = 0 \ P\text{-a.s.} \,, \tag{9}$$

where $\Theta(P)$ is understood to be a set of measurable functions from $\mathbf{R}^{d_w}$ to $\mathbf{R}$. For $1 \leq q \leq \infty$, the distribution P is said to be L^q-complete with respect to W if condition (9) holds with $\Theta(P) = L^q(P_W)$. Here, P_W denotes the marginal distribution of W under P and $L^q(P_W)$ denotes (up to P_W-equivalence classes) the set of measurable functions from $\mathbf{R}^{d_w}$ to $\mathbf{R}$ with finite $L^q(P_W)$-norm. For the special cases in which $q = 1$ or $q = \infty$, P is sometimes simply said to be complete with respect to W or bounded complete with respect to W, respectively. See d'Haultfoeuille (2011) and Andrews (2011) for further discussion.

We restrict attention to sets of measures $\mathbf{P}$ that have a common dominating measure. Specifically, letting $\mathbf{M}_{x,z}$ denote the set of all probability measures on $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, and defining

$$\mathbf{M}_{x,z}(\nu) \equiv \{P \in \mathbf{M}_{x,z} : P \ll \nu\} \,, \tag{10}$$

for some Borel measure ν on $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, we let $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$. In this setting, the null hypothesis corresponds to the set of measures for which the completeness condition fails, and hence we let

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : (9) \text{ holds under } P\} \,. \tag{11}$$

Remark 3.1. We note that the null hypothesis $H_0 : P \in \mathbf{P}_0$ states that P is such that the completeness condition does not hold. For illustrative purposes, it is helpful to draw an analogy to a simple parametric model where $(Y_i, X_i, Z_i) \in \mathbf{R}^3$ are distributed according to P, and in addition for some $\theta \in \mathbf{R}$:

$$Y_i = X_i \theta + \epsilon_i \quad \text{with} \quad E_P[Z_i(Y_i - X_i\theta)] = 0 \,. \tag{12}$$

In this context, the usual rank condition required for identification is $E_P[X_i Z_i] \neq 0$, and thus our setup is analogous to testing $H_0 : E_P[X_i Z_i] = 0$ against $H_1 : E_P[X_i Z_i] \neq 0$. $\blacksquare$

We will make use of the following assumptions in establishing our result:

Assumption 3.1. ν *is a positive* $\sigma-$*finite Borel measure on* $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$.

Assumption 3.2. $\nu = \nu_x \times \nu_z$, *where* ν_x *and* ν_z *are Borel measures on* $\mathbf{R}^{d_x}$ *and* $\mathbf{R}^{d_z}$, *respectively.*

Assumption 3.3. *The measure* ν_x *is atomless (on* $\mathbf{R}^{d_x}$).

Note that since we impose the requirement that $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for some ν satisfying Assumptions 3.1, 3.2 and 3.3, properties of ν translate into restrictions on $\mathbf{P}$. For instance, if ν has bounded support, then $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$ implies that the support of (X_i, Z_i) under P is uniformly bounded in $P \in \mathbf{P}$. In particular, by choosing ν_x and ν_z to be the Lebesgue measure on $[0,1]^{d_x}$ and $[0,1]^{d_z}$, respectively, we may impose the requirement that the support of (X_i, Z_i) under P is contained in $[0,1]^{d_x} \times [0,1]^{d_z}$ for all $P \in \mathbf{P}$. See Hall and Horowitz (2005) and Horowitz and Lee (2007) for examples of the use of such an assumption. It is also worth emphasizing that while Assumption 3.2 imposes that ν be a product measure, the requirement that $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for some such ν does not imply that each $P \in \mathbf{P}$ is itself of such form. On the other hand, the requirement that $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for some ν satisfying Assumptions 3.2 and 3.3, does imply that $P\{X_i \neq Z_i\} > 0$ for all $P \in \mathbf{P}$. Finally, we point out that if $d_x > 1$, then Assumption 3.3 may be weakened to instead requiring that at least one component of X_i have an atomless marginal measure. For ease exposition, however, we impose the stronger than necessary requirement in Assumption 3.3.

Theorem 3.1. *Suppose Assumptions 3.1, 3.2 and 3.3 hold. If* $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$, *for* $\mathbf{M}_{x,z}(\nu)$ *as in* (10), *and* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *are as in* (11) *with* $\Theta(P) = L^\infty(P_W)$, *then, for any sequence of tests* $\{\phi_n\}_{n=1}^\infty$

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \quad \text{for all } n \geq 1 \ . \tag{13}$$

Theorem 3.1 establishes the nonexistence of nontrivial tests for bounded completeness. The conclusion of Theorem 3.1 continues to hold if $\Theta(P)$ instead satisfies $L^\infty(P_W) \subseteq \Theta(P)$. Any such modification only enlarges $\mathbf{P}_0$, and hence $\mathbf{P}_0$ continues to be dense in $\mathbf{P}$ with respect to the Total Variation distance. In particular, by setting $\Theta(P) = L^q(P_W)$ for any $1 \leq q < \infty$, we are able to conclude that there exist no nontrivial test of L^q-completeness conditions either.

While we focus on the testability of L^q-completeness conditions due to their importance in the literature, it is worth noting that they are only necessary conditions for identification. Indeed, if we consider P as a measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ instead, then P_{XZ} being L^q-complete with respect to W does not guarantee existence of a solution (in $\theta \in L^q(P_W)$) to the equation

$$E_P[Y_i - \theta(W_i)|Z_i] = 0 \ P\text{-a.s.} \tag{14}$$

Rather, P_{XZ} being L^q-complete with respect to W just ensures that if a solution to (14) does exist, then it must be unique – see Proposition 2.1 in Newey and Powell (2003). Thus, P_{XZ} being L^q-complete with respect to W is only equivalent to identification under the additional assumption that (14) holds for some $\theta \in L^q(P_W)$.

Remark 3.2. In establishing Theorem 3.1, we construct for each $P \in \mathbf{P}_1$ a sequence $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ such that $\|P - P_k\|_{TV} = o(1)$. This approach requires us to exhibit for each P_k a corresponding function $\theta_k \in \Theta(P_k)$ such that $\theta_k \neq 0$ P_k-a.s. and $E_{P_k}[\theta_k(X_i)|Z_i] = 0$ P_k-a.s. While the θ_k that appear in the proof are not differentiable everywhere, it is worth emphasizing this is not an essential feature of the argument. In particular, by using Lemma 2.1 in Santos (2012), the P_k may be chosen so that each corresponding θ_k is in fact infinitely differentiable. Therefore, no nontrivial test exists even if $\Theta(P)$ is further restricted to be a smooth class of functions, such as a Hölder space. By rescaling θ_k appropriately, we may in fact even restrict $\Theta(P)$ to be a Hölder ball. ∎

Remark 3.3. As argued in Remark 2.1, it may be possible to restore testability by further restricting $\mathbf{P}$. Unfortunately, standard smoothness conditions are inadequate for this purpose. For instance, if ν is the Lebesgue measure, $\mathbf{P}$ can be restricted to be the set of $P \in \mathbf{M}_{x,z}(\nu)$ such that $dP/d\nu$ lies in a Hölder ball, and the support of $P \in \mathbf{P}$ is contained in a common compact set. A construction in Lemma 2.1 in Santos (2012), however, shows that if $dP/d\nu$ is polynomial of finite order, then P does not satisfy a completeness condition. Since polynomial densities are smooth and dense with respect to $\|\cdot\|_{TV}$ in $\mathbf{P}$, it follows that Lemma 2.1 again delivers (13). ∎

Remark 3.4. In the context of identification of some linear, semiparametric models with endogeneity, full rank requirements on matrices arise instead of completeness conditions. Specifically,

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : E_P[Z_i W_i'] \text{ has full rank}\} .$$

Tests for this purpose have been proposed, among others, by Anderson (1951), Gill and Lewbel (1992), Cragg and Donald (1993, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006). Contrary to the implications of Theorem 3.1, nontrivial tests that control asymptotic size do exist, for example, if the support of (X_i, Z_i) under P is bounded uniformly in $P \in \mathbf{P}$. ∎

Remark 3.5. Under additional restrictions, the requirement that ν_x be atomless in Assumption 3.3 may be relaxed to it being a mixture of an atomless and a discrete measure. However, the conclusion of Theorem 3.1 may not apply if ν_x is a purely discrete measure. For example, suppose that ν_x and ν_z have finite support $\{x_1, \ldots, x_s\}$ and $\{z_1, \ldots, z_t\}$. Let $\Pi(P)$ be the $s \times t$ matrix with entry $\Pi(P)_{jk} = P\{X_i = x_j | Z_i = z_k\}$. Theorem 2.4 in Newey and Powell (2003) fully characterizes L^1-completeness of P with respect to W in terms of rank conditions on submatrices of $\Pi(P)$. In this setting, nontrivial tests for L^1-completeness can therefore be constructed using, for example, uniform confidence regions for $\Pi(P)$. See Anderson (1967) and Romano and Wolf (2000) for relevant results about confidence regions for a univariate mean. ∎

3.2 Testing Identification

In this section we proceed to consider two widely studied nonparametric models with endogeneity that impose conditional quantile independence restrictions. Due to their nonlinear nature, simple,

global rank conditions such as completeness conditions are unavailable for these models, and it is for this reason that we directly examine tests for identification instead.

Throughout this section, we let $V_i = (Y_i, X_i, Z_i) \in \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ be random variables distributed according to $P \in \mathbf{P}$. As before, we let $Z_i = (Z_i^{(1)}, Z_i^{(2)})$, with the subvector $Z_i^{(1)}$ possibly empty, and let $W_i = (X_i, Z_i^{(1)})$. We will once again focus on sets of distributions $\mathbf{P}$ that are dominated by a common measure ν. Thus, by analogy with the notation used in the preceding section, we let $\mathbf{M}_{y,x,z}$ be the set of all probability measures on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ and define

$$\mathbf{M}_{y,x,z}(\nu) \equiv \{P \in \mathbf{M}_{y,x,z} : P \ll \nu\} \ . \tag{15}$$

We will impose the following requirements on the dominating measure ν:

Assumption 3.4. *ν is a positive σ−finite Borel measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$.*

Assumption 3.5. *$\nu = \nu_y \times \nu_x \times \nu_z$, where ν_y, ν_x and ν_z are Borel measures on $\mathbf{R}$, $\mathbf{R}^{d_x}$ and $\mathbf{R}^{d_z}$, respectively.*

Assumptions 3.4 and 3.5 are modifications of Assumptions 3.1 and 3.2 from the previous section to account for the fact that here the random variables take values in $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ rather than just $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$. Note that in the following two theorems we impose the requirement that $\mathbf{P} \subseteq \mathbf{M}_{y,x,z}(\nu)$ for some ν satisfying Assumptions 3.3, 3.4 and 3.5. As in the previous section, properties of ν therefore translate into restrictions on $\mathbf{P}$. See the discussion preceding Theorem 3.1. Finally, we note that in each of the following two theorems Assumption 3.3 may also be relaxed in the same way as described preceding Theorem 3.1.

3.2.1 Single Quantile Restriction Model

The first model we consider is one where for an outcome of interest Y_i, an endogenous variable X_i, and an instrumental variable Z_i and each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P)$ such that

$$Y_i = \theta(W_i) + \epsilon_i \text{ and } P\{Y_i - \theta(W_i) \leq 0 | Z_i\} = \tau \text{ w.p.1 under } P \tag{16}$$

for some pre-specified $\tau \in (0, 1)$. Here, $\Theta(P)$ is a set of measurable functions from $\mathbf{R}^{d_w}$ to $\mathbf{R}$, often set to equal $L^q(P_W)$ for some $1 \leq q \leq \infty$. In analogy to our analysis of the testability of completeness conditions, we let the null hypothesis be that identification fails in (16), and the alternative hypothesis be that it holds. To this end, we therefore define

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : \exists! \ \theta \in \Theta(P) \text{ s.t. (16) holds under } P\} \ , \tag{17}$$

where uniqueness of $\theta \in \Theta(P)$ is understood to be up to sets of measure zero under P.

The next result shows that no nontrivial test of identification exists in models defined by (16).

Theorem 3.2. *Suppose ν satisfies Assumptions 3.3, 3.4 and 3.5. Define $\mathbf{M}_{y,x,z}(\nu)$ as in (15) and let $\mathbf{P}$ be the set of all $P \in \mathbf{M}_{y,x,z}(\nu)$ for which there is some $\theta \in \Theta(P) = L^\infty(P_W)$ such that (16) holds. If $\mathbf{P}_0$ and $\mathbf{P}_1$ are as in (17), then for any sequence of tests $\{\phi_n\}_{n=1}^\infty$*

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \quad \text{for all } n \geq 1 . \tag{18}$$

In establishing Theorem 3.2, we show that for every $P \in \mathbf{P}_1$ there exists a sequence $\{P_k\}_{k=1}^\infty$ in $\mathbf{P}_0$ such that $\|P - P_k\|_{TV} = o(1)$. Our construction does not exploit the fact that $P \in \mathbf{P}_1$, but rather just the fact that $P \in \mathbf{M}_{y,x,z}(\nu)$. It therefore follows that $\mathbf{P}_0$ is actually dense in $\mathbf{M}_{y,x,z}(\nu)$ with respect to the Total Variation distance. As a result, the conclusion of Theorem 3.2 continues to hold if we instead set $\Theta(P) = L^q(P_W)$ for any $1 \leq q < \infty$. It is worth noting that, in contrast to the setting of Theorem 3.1, here letting $\Theta(P) = L^q(P_W)$ for $1 \leq q < \infty$ enlarges $\mathbf{P}$ itself, and so potentially enlarges not only $\mathbf{P}_0$, but also $\mathbf{P}_1$.

Remark 3.6. In establishing the denseness of $\mathbf{P}_0$ in $\mathbf{P}$, we construct a sequence $\{P_k\}_{k=1}^\infty$ in $\mathbf{P}_0$ such that for each k there exist functions $\theta_k^{(1)}$ and $\theta_k^{(2)}$ in $\Theta(P_k)$ that differ not only on a set with positive probability under P_k, but in the stronger sense of

$$E_{P_k}[(1\{Y_i \leq \theta_k^{(1)}(X_i)\} - 1\{Y_i \leq \theta_k^{(2)}(X_i)\})^2] > 0 ,$$

while still satisfying

$$P_k\{Y_i \leq \theta_k^{(1)}(X_i)|Z_i\} = P_k\{Y_i \leq \theta_k^{(2)}(X_i)|Z_i\} = \tau$$

w.p.1 under P_k. This feature of the proof is noteworthy because it may still be the case that $1\{Y_i \leq \theta_k^{(1)}(X_i)\} = 1\{Y_i \leq \theta_k^{(2)}(X_i)\}$ w.p.1 under P_k for functions $\theta_k^{(1)}$ and $\theta_k^{(2)}$ that differ with positive probability under P_k. ■

3.2.2 Nonseparable Model

The final model we consider is closely related to the single quantile independence model in (16). Specifically, for an outcome of interest Y_i, an endogenous variable X_i, and an instrumental variable Z_i, we now consider a setting in which for each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P)$ such that

$$Y_i = \theta(W_i, \epsilon_i) \text{ and } P\{Y_i - \theta(W_i, \tau) \leq 0|Z_i\} = \tau \text{ w.p.1 under } P \text{ for all } \tau \in (0, 1) . \tag{19}$$

Here, $\Theta(P)$ denotes a set of measurable functions $\theta : \mathbf{R}^{d_w} \times [0, 1] \to \mathbf{R}$ such that $\theta(W_i, \cdot)$ is strictly increasing P-a.s. Often, boundedness restrictions are imposed on θ, and $\Theta(P)$ is set to equal

$$\mathbf{T}(P) \equiv \left\{ \theta \in \mathbf{T} : \theta(W_i, \cdot) \text{ is strictly increasing } P\text{-a.s. and } \sup_{0 \leq \tau \leq 1} \|\theta(\cdot, \tau)\|_{L^\infty(P)} < \infty \right\} , \tag{20}$$

where $\mathbf{T}$ denotes the set of all measurable functions $\theta : \mathbf{R}^{d_w} \times [0, 1] \to \mathbf{R}$.

We again focus on the problem of testing directly for identification. To this end, we define

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : \exists!\ \theta \in \Theta(P) \text{ s.t. (19) holds under } P\} , \tag{21}$$

where uniqueness of $\theta \in \Theta(P)$ is again understood to be up to sets of measure zero under P. It is worth noting that if (19) holds for a single fixed $\tau \in (0,1)$, then this model is equivalent to the model in (16) as we can always re-write $Y_i = \theta(W_i, \tau) + \tilde{\epsilon}_i$ for $\tilde{\epsilon}_i \equiv \theta(W_i, \epsilon_i) - \theta(W_i, \tau)$. As a result of this connection between models (16) and (19), it is perhaps to be expected that the conclusion of Theorem 3.2 extends to the present setting. The following result establishes this point, showing that under commonly used restrictions for $\mathbf{P}$, no nontrivial test of identification exists in models defined by (19) even if we impose that $\theta(W_i, \cdot)$ be strictly increasing w.p.1 under P.

Theorem 3.3. *Let Assumptions 3.3, 3.4 and 3.5 hold, and let* $\mathbf{M}_{y,x,z}(\nu)$ *be as in (15) and* $\mathbf{T}(P)$ *be as in (20). If* $\mathbf{P}$ *equals the set of all* $P \in \mathbf{M}_{y,x,z}(\nu)$ *for which there is some* $\theta \in \Theta(P) = \mathbf{T}(P)$ *such that (19) holds, and* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *are as in (21), then for any sequence of tests* $\{\phi_n\}_{n=1}^{\infty}$

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \ \text{ for all } n \geq 1 . \tag{22}$$

As in Theorem 3.2, our proof implies that $\mathbf{P}_0$ is dense in $\mathbf{M}_{y,x,z}(\nu)$ with respect to the Total Variation metric. It therefore follows that the conclusion of Theorem 3.3 continues to hold if we instead require that each $\theta \in \Theta(P)$ be such that $\theta(W_i, \cdot)$ be strictly increasing and $\|\theta(\cdot, \tau)\|_{L^q(P)} < \infty$ for all $\tau \in (0,1)$ and any $1 \leq q < \infty$. Moreover, denseness of $\mathbf{P}_0$ is established by constructing sequences $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ such that for each k there exist $\theta_k^{(1)}$ and $\theta_k^{(2)}$ in $\Theta(P_k)$ satisfying

$$E_{P_k}[(1\{Y_i \leq \theta_k^{(1)}(X_i, \tau)\} - 1\{Y_i \leq \theta_k^{(2)}(X_i, \tau)\})^2] > 0$$

for all $\tau \in (0,1)$. Thus, $\theta_k^{(1)}(\cdot, \tau)$ and $\theta_k^{(2)}(\cdot, \tau)$ differ for every τ not just on a set with positive probability under P_k, but in the stronger sense of Remark 3.6.

4   Conclusion

We have provided conditions under which nontrivial tests do not exist for completeness conditions or identification in some nonparametric models with endogeneity. Our results should not be interpreted as an indictment against nonparametric methods in models with endogeneity. As any other nontestable assumption, they can still be appropriate in applications where an empirical researcher has other reasons for believing them to be true. Alternatively, inferential methods have also been developed that do not rely on such assumptions by allowing for partial identification.

Whether testability of these assumptions can be restored under stronger requirements on either θ or $\mathbf{P}$ remains an important open question. Unfortunately, Remarks 3.2 and 3.3 suggest that standard smoothness restrictions are insufficient for this purpose. We hope the results and arguments in this paper provide some guidance for addressing these challenges in future research.

5 Appendix

Throughout the Appendix we employ the following notation, not necessarily introduced in the text.

$A \triangle B$ For two sets A and B, $A \triangle B \equiv (A \setminus B) \cup (B \setminus A)$.

$\mathcal{B}(A)$ For a Borel set A, the σ-algebra generated by all open subsets of A.

$\| \cdot \|_{L^q(\lambda)}$ For $1 \le q < \infty$, a measure λ, and measurable function f, $\|f\|_{L^q(\lambda)}^q \equiv \int |f(u)|^q \lambda(du)$.

$\| \cdot \|_{L^\infty(\lambda)}$ For a measure λ, and measurable function f, $\|f\|_\infty \equiv \inf\{M > 0 : |f(u)| \le M \text{ for } \lambda\text{-a.s.}\}$.

$L^q(\lambda)$ For $1 \le q \le \infty$ and a measure λ, $L^q(\lambda)$ is the space of (λ-equivalence classes) of measurable functions f such that $\|f\|_{L^q(\lambda)} < \infty$

Lemma A.1. *Let $A \subseteq \mathbf{R}^d$ be a Borel set, $\mathcal{B}(A)$ the σ-algebra generated by all open subsets of A, and λ an atomless positive Borel measure on $\mathbf{R}^d$ satisfying $0 < \lambda\{A\} < \infty$. Then, there exists a map $\bar{B} : [0,1] \to \mathcal{B}(A)$ such that: (i) $\bar{B}(0) = \emptyset$ and $\bar{B}(1) = A$, (ii) $\bar{B}(\tau) \subseteq \bar{B}(\tau')$ for all $0 \le \tau \le \tau' \le 1$, (iii) $\lambda\{\bar{B}(\tau)\} = \tau\lambda\{A\}$. Additionally, there is $\tilde{B} : [0,1] \to \mathcal{B}(A)$ satisfying properties (i)-(iii) and such that $\lambda\{\bar{B}(\tau)\triangle\tilde{B}(\tau)\} > 0$ for all $\tau \in (0,1)$.*

PROOF: Without loss of generality we assume $\lambda\{A\} = 1$, otherwise we may just renormalize. Let μ denote the Lebesgue measure, and $\mathcal{B}([0,1])$ the σ-algebra generated by all open subsets of $[0,1]$. For any $U_1, U_2 \in \mathcal{B}(A)$, define the equivalence relation $U_1 \sim U_2$ if $\lambda\{U_1\triangle U_2\} = 0$, and denote the set of resulting equivalence classes by E_λ. Similarly, denote by E_μ the equivalence classes on $\mathcal{B}([0,1])$ generated by μ. Next, observe that since λ is a Borel measure on $\mathbf{R}^d$, Theorem 7.1.7 in Bogachev (2007b) implies λ is Radon, and hence also separable by Proposition 7.14.12(ii) in Bogachev (2007b). It therefore follows from Theorem 9.3.4 in Bogachev (2007b) that (E_λ, λ) is isomorphic to (E_μ, μ), i.e., there exists a one to one mapping $\Gamma : E_\lambda \to E_\mu$ such that:

$$\mu\{\Gamma(U_1)\} = \lambda\{U_1\}, \quad \Gamma(U_1 \setminus U_2) = \Gamma(U_1) \setminus \Gamma(U_2), \quad \Gamma(U_1 \cup U_2) = \Gamma(U_1) \cup \Gamma(U_2) \qquad (23)$$

for any $U_1, U_2 \in E_\lambda$. Next, define a map $B : (0,1) \to \mathcal{B}(A)$ satisfying $B(\tau) \in \Gamma^{-1}([0,\tau])$ for any $\tau \in (0,1)$, and finally let $\bar{B} : [0,1] \to \mathcal{B}(A)$ be given by $\bar{B}(0) = \emptyset$, $\bar{B}(1) = A$ and for any $\tau \in (0,1)$:

$$\bar{B}(\tau) \equiv [B(\tau) \cup \{ \bigcup_{0 \le \sigma < \tau, \sigma \in \mathbf{Q}} B(\sigma)\}] \cap \{ \bigcap_{1 \ge \sigma > \tau, \sigma \in \mathbf{Q}} B(\sigma)\} , \qquad (24)$$

where $\mathbf{Q}$ are the rational numbers. By construction, $\bar{B}$ then satisfies properties (i) and (ii). Moreover, for any $\tau \in (0,1)$, let $\{a_i(\tau)\}_{i=1}^\infty = \mathbf{Q} \cap [0,\tau)$, and note that by (23) we have:

$$\lambda\{\bar{B}(\tau) \setminus B(\tau)\} \le \lim_{n \to \infty} \lambda\{\{\bigcup_{i=1}^n B(a_i(\tau))\} \setminus B(\tau)\} = \lim_{n \to \infty} \mu\{\{\bigcup_{i=1}^n [0, a_i(\tau)]\} \setminus [0,\tau]\} = 0 . \qquad (25)$$

Similarly, letting $\{b_i(\tau)\}_{i=1}^\infty = \mathbf{Q} \cap (\tau, 1]$ we also obtain by monotone convergence that:

$$\lambda\{B(\tau) \setminus \bar{B}(\tau)\} = \lim_{n \to \infty} \lambda\{B(\tau) \setminus \{\bigcap_{i=1}^n B(b_i(\tau))\}\} = \lim_{n \to \infty} \mu\{[0,\tau] \setminus \{\bigcap_{i=1}^n [0, b_i(\tau)]\}\} = 0 . \qquad (26)$$

We conclude from (25) and (26) that $\lambda\{\bar{B}(\tau)\} = \lambda\{B(\tau)\}$, and since $\lambda\{B(\tau)\} = \mu\{[0,\tau]\} = \tau$ due to (23) it follows that $\bar{B}$ satisfies property (iii) as well.[1]

In order to establish the second claim of the Lemma, pointwise define $\tilde{B}: [0,1] \to \mathcal{B}(A)$ by:

$$\tilde{B}(\tau) \equiv A \setminus \bar{B}(1 - \tau) . \tag{27}$$

It is then immediate that $\tilde{B}(0) = \emptyset$ and $\tilde{B}(1) = A$, while $\lambda\{\bar{B}(\tau)\} = \tau$ additionally yields that:

$$\lambda\{\tilde{B}(\tau)\} = \lambda\{A \setminus \bar{B}(1 - \tau)\} = 1 - (1 - \tau) = \tau . \tag{28}$$

Furthermore, for any $0 \le \tau \le \tau' \le 1$, note that $\tau \le \tau'$ implies $\bar{B}(1 - \tau') \subseteq \bar{B}(1 - \tau)$, and therefore:

$$\tilde{B}(\tau) = A \setminus \bar{B}(1 - \tau) \subseteq A \setminus \bar{B}(1 - \tau') = \tilde{B}(\tau') . \tag{29}$$

Thus, from (28) and (29) we obtain that $\tilde{B}: [0,1] \to \mathcal{B}(A)$ indeed satisfies properties (i)-(iii). To conclude, note monotonicity of $\bar{B}$ implies $(A \setminus \bar{B}(1 - \tau)) \setminus \bar{B}(\tau) = A \setminus \bar{B}(\max\{\tau, 1 - \tau\})$, and hence:

$$\lambda\{\tilde{B}(\tau) \triangle \bar{B}(\tau)\} \ge \lambda\{\tilde{B}(\tau) \setminus \bar{B}(\tau)\} = \lambda\{A \setminus \bar{B}(\max\{\tau, 1 - \tau\})\} = (1 - \max\{\tau, 1 - \tau\}) . \tag{30}$$

Therefore, it follows from (30) that $\lambda\{\tilde{B}(\tau) \triangle \bar{B}(\tau)\} > 0$ for all $\tau \in (0,1)$. ∎

Lemma A.2. *Let Q be a Borel probability measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ satisfying $Q \ll \lambda$ for λ a σ-finite positive Borel measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$. If $f \equiv dQ/d\lambda$, then there exists a sequence $\{f_n\}_{n=1}^{\infty}$ satisfying: (i) $f_n \ge 0$, (ii) $\int f_n d\lambda = 1$, (iii) $\|f_n - f\|_{L^1(\lambda)} = o(1)$, and (iv) of the form:*

$$f_n(y, x, z) = \sum_{1 \le i, j, l \le K_n} \pi_{ijln} 1\{(y, x, z) \in S_{ijln}\} , \tag{31}$$

where $S_{ijln} = A_{in} \times B_{jn} \times C_{ln}$ and for some $M_n > 0$ the collections $\{A_{in}\}_{i=1}^{K_n}$, $\{B_{jn}\}_{j=1}^{K_n}$, $\{C_{ln}\}_{l=1}^{K_n}$ are partitions of $[-M_n, M_n]$, $[-M_n, M_n]^{d_x}$ and $[-M_n, M_n]^{d_z}$ respectively.

PROOF: Note $f \ge 0$ and $f \in L^1(\lambda)$. Since λ is a Borel measure it is also regular by Theorem 7.1.7 in Bogachev (2007b), and hence Theorem 13.9 in Aliprantis and Border (2006) implies there is a sequence $\{f_n^\star\}_{n=1}^{\infty}$ of continuous, compactly supported functions such that $f_n^\star \ge 0$ for all n and:

$$\|f_n^\star - f\|_{L^1(\lambda)} = o(1) . \tag{32}$$

Next, let $\Omega_n \subset \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ be the compact support of $f_n^\star$ and select $M_n > 0$ sufficiently large so that $\Omega_n \subseteq [-M_n, M_n]^{1+d_x+d_z}$. Additionally, select $\xi_n \downarrow 0$ such that $\lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(\xi_n^{-1})$, and notice that $f_n^\star$ being uniformly continuous on $[-M_n, M_n]^{1+d_x+d_z}$ implies there exist partitions $\{A_{in}\}_{i=1}^{K_n}$, $\{B_{jn}\}_{j=1}^{K_n}$ and $\{C_{ln}\}_{l=1}^{K_n}$ of $[-M_n,, M_n]$, $[-M_n, M_n]^{d_x}$ and $[-M_n, M_n]^{d_z}$, such that:

$$\max_{1 \le i, j, l \le K_n} \sup_{(y, x, z) \in A_{in} \times B_{jn} \times C_{ln}} \sup_{(\tilde{y}, \tilde{x}, \tilde{z}) \in A_{in} \times B_{jn} \times C_{ln}} |f_n^\star(y, x, z) - f_n^\star(\tilde{y}, \tilde{x}, \tilde{z})| \le \xi_n . \tag{33}$$

---

[1]We would like to thank an anonymous referee for suggesting this method of proof.

Letting $S_{ijln} \equiv A_{in} \times B_{jn} \times C_{ln}$ for $1 \leq i, j, l \leq K_n$, we then pointwise define the functions:

$$\tilde{f}_n(y, x, z) \equiv \sum_{1 \leq i,j,l \leq K_n} \tilde{\pi}_{ijln} 1\{(y, x, z) \in S_{ijln}\} \qquad \tilde{\pi}_{ijln} \equiv \sup_{(y,x,z) \in S_{ijln}} f_n^\star(y, x, z) . \qquad (34)$$

By construction, $\tilde{\pi}_{ijln} \geq 0$ for all $1 \leq i, j, l \leq K_n$ and all n, and hence $\tilde{f}_n \geq 0$. Moreover, since $f_n^\star, \tilde{f}_n$ vanish outside $[-M_n, M_n]^{1+d_x+d_z}$, and $\{S_{ijln}\}_{1 \leq i,j,l \leq K_n}$ is a partition of $[-M_n, M_n]^{1+d_x+d_z}$, equations (33) and (34) imply $\sup_{(y,x,z)} |f_n^\star(y, x, z) - \tilde{f}_n(y, x, z)| \leq \xi_n$ λ-a.s. Thus, we obtain:

$$\|f_n^\star - \tilde{f}_n\|_{L^1(\lambda)} = \int_{[-M_n,M_n]^{1+d_x+d_z}} |f_n^\star - \tilde{f}_n| d\lambda \leq \xi_n \lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(1) , \qquad (35)$$

since $\xi_n \downarrow 0$ was chosen so that $\lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(\xi_n^{-1})$. Finally, let $f_n \equiv \tilde{f}_n / \|\tilde{f}_n\|_{L^1(\lambda)}$ and note properties (i) and (ii) follow while (31) holds with $\pi_{ijln} = \tilde{\pi}_{ijln} / \|\tilde{f}_n\|_{L^1(\lambda)}$. Moreover, since $\|\tilde{f}_n - f\|_{L^1(\lambda)} = o(1)$ by (32) and (35) it follows that $\|\tilde{f}_n\|_{L^1(\lambda)} \to 1$, and we conclude that:

$$\|f_n - f\|_{L^1(\lambda)} \leq \frac{\|\tilde{f}_n - f\|_{L^1(\lambda)}}{\|\tilde{f}_n\|_{L^1(\lambda)}} + |1 - \frac{1}{\|\tilde{f}_n\|_{L^1(\lambda)}}| \times \|f\|_{L^1(\lambda)} = o(1) , \qquad (36)$$

which verifies property (iii) and hence the claim of the Lemma follows. ∎

PROOF OF LEMMA 2.1: Fix $P \in \mathbf{P}_1$ and let $\{P_k\}_{k=1}^\infty$ satisfy $\|P_k - P\|_{TV} = o(1)$ and $P_k \in \mathbf{P}_0$ for all k. Since for any sequence of scalars $\{a_i\}_{i=1}^n$, $\{b_i\}_{i=1}^n$ satisfying $|a_i| \leq 1$ and $|b_i| \leq 1$ for all $1 \leq i \leq n$ we have $|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i| \leq \sum_{i=1}^n |a_i - b_i|$, it follows that:

$$\lim_{k \to \infty} \sup_{A \in \mathcal{B}^n(\mathbf{V})} |P_k^n(A) - P^n(A)| \leq n \times \lim_{k \to \infty} \sup_{A \in \mathcal{B}(\mathbf{V})} |P_k(A) - P(A)| = 0 , \qquad (37)$$

where $\mathcal{B}(\mathbf{V})$ denotes the $\sigma-$algebra generated by all open sets of $\mathbf{V}$, and $\mathcal{B}^n(\mathbf{V}) \equiv \bigotimes_{i=1}^n \mathcal{B}(\mathbf{V})$. Since result (37) implies that for every n, $\|P_k^n - P^n\|_{TV} = o(1)$ as $k \uparrow \infty$, the Lemma follows from applying Theorem 1 in Romano (2004) to the sets $\mathbf{P}_{0,n} \equiv \{Q \in \bigotimes_{i=1}^n \mathbf{M} : Q = P^n \text{ some } P \in \mathbf{P}_0\}$ and $\mathbf{P}_{1,n} \equiv \{Q \in \bigotimes_{i=1}^n \mathbf{M} : Q = P^n \text{ some } P \in \mathbf{P}_1\}$. ∎

PROOF OF THEOREM 3.1: Fix $P \in \mathbf{P}_1$ and let $f \equiv dP/d\nu$. By Assumption 3.1 and Lemma A.2 applied to $Q \equiv \delta_0 \times P$ and $\lambda \equiv \delta_0 \times \nu$ for δ_0 a degenerate measure at 0 on $\mathbf{R}$, there is $\{f_k\}_{k=1}^\infty$ with:

$$\|f_k - f\|_{L^1(\nu)} = o(1) , \qquad (38)$$

and in addition, for all k each f_k satisfies $f_k \geq 0$, $\int f_k d\nu = 1$, and is a simple function of the form:

$$f_k(x, z) = \sum_{1 \leq j,l \leq K_k} \pi_{jlk} 1\{(x, z) \in S_{jlk}\} . \qquad (39)$$

Here, $S_{jlk} = B_{jk} \times C_{lk}$ and for some $M_k > 0$, the collections $\{B_{jk}\}_{j=1}^{K_k}$ and $\{C_{lk}\}_{l=1}^{K_k}$ are partitions of $[-M_k, M_k]^{d_x}$ and $[-M_k, M_k]^{d_z}$ respectively. Therefore, defining:

$$P_k\{E\} \equiv \int_E f_k d\nu \qquad (40)$$

14

for all Borel measurable $E \subseteq \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, it follows that P_k is a probability measure with $P_k \ll \nu$, and hence $P_k \in \mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for all k. Moreover, by (38), $\{P_k\}_{k=1}^\infty$ satisfies $\|P - P_k\|_{TV} = o(1)$.

In what follows, we aim to show that in fact $P_k \in \mathbf{P}_0$ for all k. Assumption 3.3 and Corollary 1.12.10 in Bogachev (2007a), then imply that for each $1 \leq j \leq K_k$ there exist Borel measurable subsets $(B_{jk}^{(1)}, B_{jk}^{(2)})$ such that $B_{jk} = B_{jk}^{(1)} \cup B_{jk}^{(2)}$, $B_{jk}^{(1)} \cap B_{jk}^{(2)} = \emptyset$ and in addition satisfy:

$$\nu_x\{B_{jk}^{(1)}\} = \nu_x\{B_{jk}^{(2)}\} = \frac{1}{2}\nu_x\{B_{jk}\} \ . \tag{41}$$

Since $\{B_{jk}\}_{j=1}^{K_k}$ is a partition of $[-M_k, M_k]^{d_x}$ by Lemma A.2, we may define a function θ_k by:

$$\theta_k(x) \equiv \sum_{j=1}^{K_k}(1\{x \in B_{jk}^{(1)}\} - 1\{x \in B_{jk}^{(2)}\}) \ . \tag{42}$$

Note that $\theta_k \in L^\infty(P_k)$, and $\theta_k \neq 0$ P_k-a.s. due to (39), (42) and $B_{jk}^{(1)} \cap B_{jk}^{(2)} = \emptyset$ for all $1 \leq j \leq K_k$. Moreover, for any bounded $z \mapsto \psi(z)$ we obtain from (39), $f_k = dP_k/d\nu$ and Assumption 3.2:

$$E_{P_k}[\psi(Z_i)\theta_k(X_i)] = \sum_{1 \leq j,l \leq K_k} \pi_{jlk} \int_{C_{lk}} \int_{B_{jk}} \psi(z)\theta_k(x)\nu_x(dx)\nu_z(dz)$$

$$= \sum_{1 \leq j,l \leq K_k} \pi_{jlk}(\nu_x\{B_{jk}^{(1)}\} - \nu_x\{B_{jk}^{(2)}\}) \int_{C_{lk}} \psi(z)\nu_z(dz) = 0 \tag{43}$$

where the final equality exploited (41). In particular, (43) must hold for $\psi(\cdot) = E_{P_k}[\theta_k(X_i)|Z_i = \cdot]$, and hence we obtain by the law of iterated expectations that $E_{P_k}[\theta_k(X_i)|Z_i] = 0$, P_k-a.s. Note that the functions $\{\theta_k\}_{k=1}^\infty$ can be viewed as a function of $W = (X, Z^{(1)})$ that only depends on X, which together with (43) suffices for concluding that $P_k \in \mathbf{P}_0$ for all k. Hence, since $P \in \mathbf{P}_1$ was arbitrary and $\|P - P_k\|_{TV} = o(1)$, the conclusion of the Theorem follows by Lemma 2.1. ∎

PROOF OF THEOREM 3.2: Fix $P \in \mathbf{P}_1$ and let $f \equiv dP/d\nu$. As in the proof of Theorem 3.1, we begin by noting that Lemma A.2 and Assumption 3.4 imply there is a sequence $\{f_k\}_{k=1}^\infty$ such that:

$$\|f_k - f\|_{L^1(\nu)} = o(1) \ , \tag{44}$$

and in addition, for all k each f_k satisfies $f_k \geq 0$, $\int f_k d\nu = 1$, and is a simple function of the form:

$$f_k(y, x, z) = \sum_{1 \leq i,j,l \leq K_k} \pi_{ijlk} 1\{(y, x, z) \in S_{ijlk}\} \ . \tag{45}$$

Here, $S_{ijlk} = A_{ik} \times B_{jk} \times C_{lk}$ and for some $M_k > 0$, $\{A_{ik}\}_{i=1}^{K_k}$, $\{B_{jk}\}_{j=1}^{K_k}$, and $\{C_{lk}\}_{l=1}^{K_k}$ form partitions of $[-M_k, M_k]$, $[-M_k, M_k]^{d_x}$, and $[-M_k, M_k]^{d_z}$ respectively. Hence, defining

$$P_k\{E\} \equiv \int_E f_k d\nu \tag{46}$$

for all Borel measurable $E \subseteq \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ we obtain a sequence of probability measure $\{P_k\}_{k=1}^\infty$ satisfying $P_k \in \mathbf{M}_{y,x,z}(\nu)$ for all k, and $\|P - P_k\|_{TV} = o(1)$ due to result (44).

We next aim to show that $P_k \in \mathbf{P}_0$ for all k. Toward this end, note that Assumption 3.3 and Lemma A.1 imply there exist collections $\{B_{jk}^{(1)}(\tau), B_{jk}^{(2)}(\tau)\}_{j=1}^{K_k}$ such that for all $1 \leq j \leq K_k$:

$$\nu_x\{B_{jk}^{(1)}(\tau)\} = \tau\nu_x\{B_{jk}\} \qquad \nu_x\{B_{jk}^{(2)}(\tau)\} = \tau\nu_x\{B_{jk}\} , \tag{47}$$

with $B_{jk}^{(l)}(\tau) \subseteq B_{jk}$ for $l \in \{1,2\}$, and $\nu_x\{B_{jk}^{(1)}(\tau)\triangle B_{jk}^{(2)}(\tau)\} > 0$ for all $1 \leq j \leq K_k$ such that $\nu_x\{B_{jk}\} > 0$. For $l \in \{1,2\}$ we may then define functions $\theta_k^{(l)}(\cdot,\tau)$ pointwise in x by:

$$\theta_k^{(l)}(x,\tau) \equiv \sum_{j=1}^{K_k}(2M_k 1\{x \in B_{jk}^{(l)}(\tau)\} - 2M_k 1\{x \in B_{jk} \setminus B_{jk}^{(l)}(\tau)\}) , \tag{48}$$

and note that $\theta_k^{(l)}(\cdot,\tau) \in L^\infty(P_k)$ for $l \in \{1,2\}$. Moreover, $Y_i \in [-M_k, M_k]$ P_k-a.s., together with Assumption 3.5 and $f_k = dP_k/d\nu$ with f_k as in (45) allows us to conclude that:

$$
\begin{aligned}
E_{P_k}&[(1\{Y_i \leq \theta_k^{(1)}(X_i,\tau)\} - 1\{Y_i \leq \theta_k^{(2)}(X_i,\tau)\})^2] \\
&= E_{P_k}[(1\{\theta_k^{(1)}(X_i,\tau) = 2M_k\} - 1\{\theta_k^{(2)}(X_i,\tau) = 2M_k\})^2] \\
&= \sum_{1 \leq i,j,l \leq K_k} \pi_{ijlk}\nu_y\{A_{ik}\}\nu_z\{C_{lk}\} \int_{B_{jk}}(1\{\theta_k^{(1)}(x,\tau) = 2M_k\} - 1\{\theta_k^{(2)}(x,\tau) = 2M_k\})^2\nu_x(dx) \\
&> 0 ,
\end{aligned}
\tag{49}
$$

where we exploited $(1\{\theta_k^{(1)}(x,\tau) = 2M_k\} - 1\{\theta_k^{(2)}(x,\tau) = 2M_k\})^2 = 1\{x \in B_{jk}^{(1)}(\tau)\triangle B_{jk}^{(2)}(\tau)\}$ for every $x \in B_{jk}$ and $\nu_x\{B_{jk}^{(1)}(\tau)\triangle B_{jk}^{(2)}(\tau)\} > 0$ for some $1 \leq j \leq K_k$ due to P_k having support $[-M_k, M_k]^{1+d_x+d_z}$ implying $P_k\{S_{ijlk}\} = \pi_{ijlk}\nu\{S_{ijlk}\} > 0$ for some $1 \leq i,j,l \leq K_k$.

We conclude from (49) that $\theta_k^{(1)}(\cdot,\tau)$ and $\theta_k^{(2)}(\cdot,\tau)$ are distinct under $\|\cdot\|_{L^\infty(P_k)}$. Additionally, for $l \in \{1,2\}$, we have $1\{\theta_k^{(l)}(x,\tau) = 2M_k\} = 1\{x \in B_{jk}^{(l)}(\tau)\}$ for every $x \in B_{jk}$ by (48), and hence $\nu_x\{\{\theta_k^{(l)}(x,\tau) = 2M_k\} \cap B_{jk}\} = \nu_x\{B_{jk}^{(l)}(\tau)\} = \tau\nu_x\{B_{jk}\}$. It follows that for any bounded $z \mapsto \psi(z)$:

$$
\begin{aligned}
E_{P_k}&[\psi(Z_i)(1\{Y_i \leq \theta_k^{(l)}(X_i,\tau)\} - \tau)] \\
&= \sum_{1 \leq i,j,l \leq K_k} \pi_{ijlk} \int_{A_{ik}}\int_{C_{lk}}\int_{B_{jk}} \psi(z)(1\{\theta_k^{(l)}(x,\tau) = 2M_k\} - \tau)\nu_x(dx)\nu_z(dz)\nu_y(dy) \\
&= \sum_{1 \leq i,j,l \leq K_k} \pi_{ijlk}(\tau\nu_x\{B_{jk}\} - \tau\nu_x\{B_{jk}\})\int_{A_{ik}}\int_{C_{lk}}\psi(z)\nu_z(dz)\nu_y(dy) \\
&= 0 .
\end{aligned}
\tag{50}
$$

In particular, setting $\psi(\cdot) = P_k\{Y_i \leq \theta^{(l)}(X_i,\tau)|Z_i = \cdot\} - \tau$ in (50), implies by the law of iterated expectations that $P_k\{Y_i \leq \theta_k^{(l)}(X_i,\tau)|Z_i\} = \tau$, P_k-a.s. for $l \in \{1,2\}$. Interpreting the functions $\{\theta_k^{(l)}\}_{k=1}^\infty$ as functions of $W = (X, Z^{(1)})$ that only depend on X, it follows that results (49) and (50) suffice for concluding that $P_k \in \mathbf{P}_0$ for all k. Hence, since $P \in \mathbf{P}_1$ was arbitrary and result (44) establishes that $\|P - P_k\|_{TV} = o(1)$, the conclusion of the Theorem follows by Lemma 2.1. ∎

PROOF OF THEOREM 3.3: The proof is very similar to that of Theorem 3.2, and we therefore provide only an outline, emphasizing the differences in the arguments. Fixing $P \in \mathbf{P}_1$, we may

obtain a sequence $\{P_k\}_{k=1}^\infty$ such that for all k, $P_k \in \mathbf{M}_{y,x,z}(\nu)$, $dP_k/d\nu = f_k$ for f_k as defined in (45), and such that (44) holds. To show $P_k \in \mathbf{P}_0$ for all k, let $\mathcal{B}(B_{jk})$ denote the σ-algebra generated by all the open subsets of B_{jk}. By Assumption 3.3 and Lemma A.1, there then exist $B_{jk}^{(1)} : [0,1] \to \mathcal{B}(B_{jk})$ and $B_{jk}^{(2)} : [0,1] \to \mathcal{B}(B_{jk})$ such that for all $1 \le j \le K_k$ with $\nu_x\{B_{jk}\} > 0$ and $l \in \{1,2\}$: (i) $\nu_x\{B_{jk}^{(l)}(\tau)\} = \tau\nu_x\{B_{jk}\}$, (ii) $B_{jk}^{(l)}(\tau) \subseteq B_{jk}^{(l)}(\tau')$ for all $0 \le \tau \le \tau' \le 1$, and (iii) $\nu_x\{B_{jk}^{(1)}(\tau) \triangle B_{jk}^{(2)}(\tau)\} > 0$ for all $\tau \in (0,1)$. Following (48), we define functions $\theta_k^{(l)}$ pointwise by:

$$\theta_k^{(l)}(x,\tau) \equiv \sum_{j=1}^{K_k} ((2+\tau)M_k 1\{x \in B_{jk}^{(l)}(\tau)\} - (3-\tau)M_k 1\{x \in B_{jk} \setminus B_{jk}^{(l)}(\tau)\}) . \tag{51}$$

Observe that $|\theta_k^{(l)}(x,\tau)| \le 3M_k$ for all $(x,\tau) \in \mathbf{R}^{d_x} \times [0,1]$, and hence $\theta_k(X_i,\tau)$ is bounded P_k-a.s. uniformly in $\tau \in [0,1]$. Moreover, since $B_{jk}^{(l)}(\tau) \subseteq B_{jk}^{(l)}(\tau')$ for $l \in \{1,2\}$ and all $0 \le \tau \le \tau' \le 1$ and $1 \le j \le K_k$, it follows from (51) and the support of X_i under P_k being contained in $[-M_k, M_k]^{d_x} = \bigcup_{j=1}^{K_k} B_{jk}$ that $\theta_k^{(l)}(X_i,\tau)$ is strictly monotonic in τ P_k-a.s. In turn, notice that since $\tau \in [0,1]$ and the support of Y_i is contained in $[-M_k, M_k]$ under P_k, we obtain from (51) that for all $x \in B_{jk}$, $1\{Y_i \le \theta_k^{(l)}(x,\tau)\} = 1\{x \in B_{jk}^{(l)}(\tau)\}$ P_k-a.s. Therefore, arguing as in (49) yields that:

$$E_{P_k}[(1\{Y_i \le \theta_k^{(1)}(X_i,\tau)\} - 1\{Y_i \le \theta_k^{(2)}(X_i,\tau)\})^2] > 0 , \tag{52}$$

for all k and $\tau \in (0,1)$. Thus, we may conclude from (52) that $\theta_k^{(1)}(\cdot,\tau)$ is distinct from $\theta_k^{(2)}(\cdot,\tau)$ under $\|\cdot\|_{L^\infty(P_k)}$ for all $\tau \in (0,1)$. In turn, arguing as in (50) further implies that for all k:

$$P_k\{Y_i \le \theta_k^{(l)}(X_i,\tau)|Z_i\} = \tau \tag{53}$$

P_k-a.s. for $l \in \{1,2\}$ and all $\tau \in (0,1)$. Since we may view the functions $\{\theta_k^{(l)}(\cdot,\tau)\}_{k=1}^\infty$ as functions of $W = (X, Z^{(1)})$ that only depend on X, results (52) and (53) suffice for concluding that $P_k \in \mathbf{P}_0$ for all k. The argument can then be finished as in the proof of Theorem 3.2. ∎

References

AI, C. and CHEN, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, **71** 1795–1843.

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

ANDERSON, T. W. (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *The Annals of Mathematical Statistics*, **22** 327–351.

ANDERSON, T. W. (1967). Confidence Limits for the Expected Value of an Arbitrary Random Variable with a Continuous Distribution Function. *Bulletin of the International Statistical Institute*, **43** 249–251.

ANDREWS, D. W. K. (2011). Examples of l^2-complete and boundedly-complete distributions. Working paper, Yale University.

BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, **27** 1115–1122.

BERRY, S. T. and HAILE, P. A. (2009). Identification of a heterogeneous generalized regression model with group effects. Discussion paper no. 1732, Cowles Foundation.

BERRY, S. T. and HAILE, P. A. (2010a). Identification in differentiated products markets using market level data. Discussion paper no. 1744, Cowles Foundation.

BERRY, S. T. and HAILE, P. A. (2010b). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Discussion paper no. 1718, Cowles Foundation.

BLUNDELL, R., CHEN, X. and KRISTENSEN, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, **75** 1613–1669.

BOGACHEV, V. I. (2007a). *Measure Theory: Volume I.* Springer-Verlag, Berlin.

BOGACHEV, V. I. (2007b). *Measure Theory: Volume II.* Springer-Verlag, Berlin.

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

CHEN, X., POUZO, D. and TAMER, E. (2011). QLR inference on partially identified nonparametric conditional moment models.

CHERNOZHUKOV, V., GAGLIARDINI, P. and SCAILLET, O. (2010). Nonparametric instrumental variable estimation of structural quantile effects. Working paper, University of Lugano.

CHERNOZHUKOV, V. and HANSEN, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73** 245–261.

CRAGG, J. G. and DONALD, S. G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, **9** 222–240.

CRAGG, J. G. and DONALD, S. G. (1997). Inferring the Rank of a Matrix. *Journal of Econometrics*, **76** 223–250.

DAROLLES, S., FAN, Y., FLORENS, J. and RENAULT, E. (2011). Nonparametric instrumental regression. *Econometrica* forthcoming.

D'HAULTFOEUILLE, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, **27** 460–471.

DONOHO, D. L. (1988). One-Sided Inference About Functionals of a Density. *The Annals of Statistics*, **16** 1390–1420.

DUFOUR, J.-M. (2003). Identification, weak instruments, and sitatistical inference in econometrics. *Canadian Journal of Economics*, **36** 767–808.

FREYBERGER, J. and HOROWITZ, J. L. (2012). Identification and shape restrictions in nonparametric IV estimation.

GILL, L. and LEWBEL, A. (1992). Testing the Rank and Definiteness of Estimated Matrices with Applications to Factor, State-Space and ARMA Models. *Journal of the American Statistical Association*, **87** 766–776.

HALL, P. and HOROWITZ, J. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, **33** 2904–2929.

HODERLEIN, S., NESHEIM, L. and SIMONI, A. (2012). Semiparametric estimation of random coefficients in structural economic models. Working paper, University College London.

HOROWITZ, J. and LEE, S. (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, **75** 1191–1208.

HU, Y. and SCHENNACH, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, **76** 195–216.

IMBENS, G. W. and NEWEY, W. K. (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, **77** 1481–1512.

KLEIBERGEN, F. and PAAP, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, **133** 97–126.

LECAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

LEEB, H. and PÖTSCHER, B. M. (2008). Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators? *Econometric Theory*, **24** 338–376.

LEHMANN, E. L. and SCHEFFÉ, H. (1950). Completeness, similar regions, and unbiased estimation - part i. **10** 305–340.

LEHMANN, E. L. and SCHEFFÉ, H. (1955). Completeness, similar regions, and unbiased estimation - part ii. **15** 219–236.

MÜLLER, U. K. (2008). The Impossibility of Consistent Discrimination between I(0) and I(1) Processes. *Econometric Theory*, **24** 616–630.

NEWEY, W. and POWELL, J. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

PÖTSCHER, B. M. (2002). Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters. *Econometrica*, **70** 1035–1065.

ROBIN, J.-M. and SMITH, R. J. (2000). Tests of Rank. *Econometric Theory*, **16** 151–175.

ROMANO, J. P. (2004). On non-parametric testing, the uniform behavior of the t-test, and related problems. *Scandinavian Journal of Statistics*, **31** 567–584.

ROMANO, J. P. and WOLF, M. (2000). Finite sample nonparametric inference and large sample efficiency. *Annals of Statistics*, **28** 756–778.

SANTOS, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, **80** 213–275.

TORGOVITSKY, A. (2011). Identification of nonseparable models with general instruments. Working paper, Yale University.