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

July 8, 2011

Abstract

This paper considers three distinct hypothesis testing problems that arise in the context of identification of some nonparametric models with endogeneity. The first hypothesis testing problem 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 concern testing 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 sequence of tests that controls asymptotic size has asymptotic power no greater than size against *any* alternative. In this sense, no nontrivial tests for these hypothesis testing problems exist.

1

1 Introduction

Let $\{V_i\}_{i=1}^n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. This paper considers three distinct hypothesis testing problems that arise in the context of identification of some nonparametric models with endogeneity. As usual, each of these hypothesis testing problems may be written as

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P}_1 \ , \tag{1}$$

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 each of the three hypothesis testing problems we consider, we provide conditions on $\mathbf{P}$ under which any sequence of tests $\{\phi_n\}_{n=1}^\infty$ that controls size at level $\alpha \in (0, 1)$ in the sense that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] \leq \alpha \tag{2}$$

also satisfies

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \alpha \ , \tag{3}$$

where P^n denotes the n-fold product measure $\bigotimes_{i=1}^n P$. We thus conclude that any test of (1) that controls size will have trivial power against *all* alternatives $P \in \mathbf{P}_1$. In this sense, we establish that no nontrivial test exists for the hypothesis testing problems under consideration.

In order to describe the first hypothesis testing problem we consider, let $V_i = (X_i, Z_i)$ and $\mathbf{P}$ be a set of probability measures on $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$. For Z_i^x a (possibly empty) subvector of Z_i and $W_i = (X_i, Z_i^x) \in \mathbf{R}^{d_w}$, define

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : E_P[\theta(W_i)|Z_i] = 0 \text{ for } \theta \in \Theta(P) \Longrightarrow \theta = 0 \ P\text{-a.s.}\} \ . \tag{4}$$

Here, $\Theta(P)$ is understood to be a subset of the set of all functions from $\mathbf{R}^{d_w}$ to $\mathbf{R}$. If $\Theta(P) = L^1(P)$, then $\mathbf{P}_1$ is the subset of $\mathbf{P}$ for which $L^1(P)$-completeness holds. If, on the other hand, $\Theta(P) = L^\infty(P)$, then $\mathbf{P}_1$ is the subset of $\mathbf{P}$ for which $L^\infty(P)$-completeness or P-bounded completeness holds. See d'Haultfoeuille (2011) and Andrews (2011) for a discussion of different completeness conditions. Such assumptions are made routinely to achieve identification in a variety of nonparametric models with endogeneity involving mean independence restrictions. For example, Newey and Powell (2003) show that $L^1(P)$-completeness is necessary for identification of the model they consider. Blundell et al. (2007) instead use $L^\infty(P)$-completeness to attain identification. See also Hall and Horowitz (2005) and Darolles et al. (2011). These assumptions have been used to achieve identification in some measurement error, random coefficient and demand models as well. See, for instance, Hu and Schennach (2008), Hoderlein et al. (2010), Berry and Haile (2010a) and the references therein. Our results establish that, under commonly used restrictions for $\mathbf{P}$, no nontrivial test of these completeness conditions exists. We also contrast this conclusion with the problem of testing rank conditions that are necessary for identification in some semiparametric models with endogeneity. See Remark 2.1.

In order to describe the second hypothesis testing problem we consider, let $V_i = (Y_i, X_i, Z_i)$ and $\mathbf{P}$ be a set of distributions on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$. As before, let Z_i^x be a (possibly empty) subvector of Z_i and $W_i = (X_i, Z_i^x) \in \mathbf{R}^{d_w}$. Consider the model for an outcome of interest Y_i, an endogenous variable X_i, and an instrumental variable Z_i where for each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P)$ for which

$$Y_i = \theta(W_i) + \epsilon_i \text{ and } P\{\epsilon_i \leq 0 | Z_i\} = \tau \text{ w.p.1 under } P \tag{5}$$

for some pre-specified $\tau \in (0, 1)$. Here, $\Theta(P)$ is a subset of the set of all functions from $\mathbf{R}^{d_w}$ to $\mathbf{R}$. This model has been studied, for example, by Chernozhukov and Hansen (2005), Horowitz and Lee (2007), Chen and Pouzo (2008) and Chernozhukov et al. (2010). In this setting, it is difficult to describe necessary conditions for identification in terms of completeness conditions. We therefore focus instead on the problem of testing for identification in this model. To this end, let

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : \exists! \ \theta \in \Theta(P) \text{ s.t. (5) holds under } P\} , \tag{6}$$

where uniqueness of $\theta \in \Theta(P)$ is understood to be up to sets of measure zero under P. Our results establish that, under commonly used restrictions for $\mathbf{P}$, no nontrivial test of identification exists in models defined by (5).

The third hypothesis testing problem we consider is closely related to the one described above. As before, let $V_i = (Y_i, X_i, Z_i)$, $\mathbf{P}$ be a set of distributions on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, Z_i^x be a (possibly empty) subvector of Z_i, and $W_i = (X_i, Z_i^x) \in \mathbf{R}^{d_w}$. Consider the model for an outcome of interest Y_i, an endogenous variable X_i, and an instrumental variable Z_i where for each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P)$ for which

$$Y_i = \theta(W_i, \epsilon_i) \text{ and } P\{\theta(W_i, \epsilon_i) - \theta(W_i, \tau) \leq 0 | Z_i\} = \tau \text{ w.p.1 under } P \text{ for all } \tau \in (0, 1) . \tag{7}$$

Here, $\Theta(P)$ denotes a subset of the set of all functions $\theta : \mathbf{R}^{d_w} \times [0, 1] \to \mathbf{R}$ such that $\theta(W_i, \cdot)$ is strictly increasing w.p.1 under P. This model has been studied, for example, by Chernozhukov and Hansen (2005), Imbens and Newey (2009) and Torgovitsky (2011). See also Berry and Haile (2009, 2010b) for how these models arise in the context of generalized regression models and multinomial choice models, respectively. We again focus on the problem of testing for identification in this model. To this end, define

$$\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 = \{P \in \mathbf{P} : \exists! \ \theta \in \Theta(P) \text{ s.t. (7) holds under } P\} , \tag{8}$$

where uniqueness of $\theta \in \Theta(P)$ is again understood to be up to sets of measure zero under P. Note that for each fixed $\tau \in (0, 1)$ this model is equivalent to the model described previously. Our results establish that, under commonly used restrictions for $\mathbf{P}$, no nontrivial test of identification exists in models defined by (7) even if we impose that $\theta(W_i, \cdot)$ be strictly increasing w.p.1 under P.

The remainder of the paper is organized as follows. We begin in Section 2.1 by introducing a useful lemma that underlies our arguments. We then describe our results for the first hypothesis

testing problem in Section 2.2. The closely related second and third hypothesis testing problems are then treated in Section 2.3. We briefly conclude in Section 3.

2 Main Results

2.1 A Useful Lemma

The following lemma underlies all of our arguments. In the statement of the lemma, $H(P, P')$ denotes the Hellinger distance between probability measures P and P'.

Lemma 2.1. *Let* $\mathbf{M}$ *denote the space of Borel probability measures on a metric space* $\mathcal{A}$*. 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* $H(P, P_k) = o(1)$*, then every sequence of test functions* $\{\phi_n\}_{n=1}^{\infty}$ *satisfies*

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] . \tag{9}$$

Lemma 2.1 is a mild modification of Theorem 1 in Romano (2004). In particular, the hypothesis of the lemma has been restated in terms of Hellinger distance, as opposed to Total Variation distance, and the conclusion has been related in terms of a large-sample result, as opposed to a finite-sample result. 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 asymptotic size. Theorem 1 in Romano (2004) establishes that the appropriate topology for this purpose is that induced by the Total Variation distance. See also Donoho (1988) for related results on the construction of confidence intervals.

In each of the three hypothesis testing problems that we consider, we establish nonexistence of nontrivial tests for identification by constructing for each $P \in \mathbf{P}_1$ a sequence $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ with $H(P, P_k) = o(1)$ 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 the Hellinger distance in all three settings we examine.

2.2 Testing Completeness

In this section, we develop our results concerning the nonexistence of nontrivial tests for completeness conditions. In order to do so, we require the following notation. Let $\mathbf{M}_{x,z}$ be the set of all probability measures on $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, and, for ν a Borel measure on $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, define

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

We will make use of the following assumptions:

4

Assumption 2.1. ν *is a positive* $\sigma-$*finite Borel measure on* $\mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$.

Assumption 2.2. $\nu = \nu_x \times \nu_z$, *where* ν_x *and* ν_z *are Borel measures on* $\mathbf{R}^{d_x}$ *and* $\mathbf{R}^{d_z}$, *respectively.*

Assumption 2.3. *The measure* ν_x *is atomless (on* $\mathbf{R}^{d_x}$).

Note that in the following theorem we impose the requirement that $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for some ν satisfying Assumptions 2.1, 2.2 and 2.3. Properties of ν therefore 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 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 2.2 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 2.2 and 2.3, 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 2.3 may be weakened to instead requiring that at least one component of X_i have an atomless marginal measure. In order to ease the exposition of our results, however, we impose the stronger than necessary requirement in Assumption 2.3.

Theorem 2.1. *Suppose* ν *satisfies Assumptions 2.1, 2.2 and 2.3. Define* $\mathbf{M}_{x,z}(\nu)$ *as in* (10) *and let* $\mathbf{P} = \mathbf{M}_{x,z}(\nu)$. *Further define* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *as in* (4) *with* $\Theta(P) = L^\infty(P)$. *If a sequence of test functions* $\{\phi_n\}_{n=1}^\infty$ *satisfies* (2) *for some* $\alpha \in (0,1)$, *then it also satisfies* (3).

Theorem 2.1 establishes the nonexistence of nontrivial tests for P-bounded completeness. The conclusion of Theorem 2.1 continues to hold if $\Theta(P)$ instead satisfies $L^\infty(P) \subseteq \Theta(P)$. Any such modification only enlarges $\mathbf{P}_0$, and hence $\mathbf{P}_0$ continues to be dense in $\mathbf{P}_1$ with respect to the Hellinger distance. In particular, by setting $\Theta(P) = L^q(P)$ for any $1 \leq q < \infty$, we are able to conclude that there exist no nontrivial tests of $L^q(P)$-completeness conditions as well.

Remark 2.1. In the context of identification of some linear, semiparametric models with endogeneity, full rank requirements on certain matrices arise instead of completeness conditions. In these settings,

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

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). In contrast to the conclusion of Theorem 2.1, nontrivial tests that satisfy (2) do exist, for example, if the support of (X_i, Z_i) under P is bounded uniformly in $P \in \mathbf{P}$. ∎

Remark 2.2. In establishing Theorem 2.1, we construct for each $P \in \mathbf{P}_1$ a sequence $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ such that $H(P, P_k) = o(1)$. This approach requires us to exhibit for each P_k a corresponding function $\theta_k \in \Theta(P)$ such that $\theta_k \neq 0$ P_k-a.s. and $E_{P_k}[\theta_k(X_i)|Z_i] = 0$. 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 (2010), 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 Sobolev space. ∎

Remark 2.3. Under additional restrictions, the requirement that ν_x be atomless in Assumption 2.3 may be relaxed to it being a mixture of an atomless and a discrete measure. However, the conclusion of Theorem 2.1 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 $\{z_1, \ldots, z_t\}$, respectively. 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) establishes that P satisfies $L^1(P)$-completeness if and only if the rank of $\Pi(P)$ is s and $s \leq t$. In this setting, nontrivial tests for $L^1(P)$-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. ∎

2.3 Testing Identification

In this section, we develop our results concerning the nonexistence of nontrivial tests for identification in certain nonparametric models with endogeneity involving quantile independence restrictions. In order to do so, we require the following notation. By analogy with the notation used in the preceding section, 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

$$\mathbf{M}_{y,x,z}(\nu) \equiv \{P \in \mathbf{M}_{y,x,z} : P \ll \nu\} , \tag{11}$$

where ν is a Borel measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$. Further let $\mathbf{T}$ denote the set of all functions $\theta : \mathbf{R}^{d_w} \times [0, 1] \to \mathbf{R}$, and define

$$\mathbf{T}(P) \equiv \left\{\theta \in \mathbf{T} : \theta(W_i, \cdot) \text{ is strictly increasing and } \sup_{0 \leq \tau \leq 1} \|\theta(\cdot, \tau)\|_{L^{\infty}(P)} < \infty\right\} . \tag{12}$$

We will make use of the following assumptions:

Assumption 2.4. ν *is a positive* $\sigma-$*finite Borel measure on* $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$.

Assumption 2.5. $\nu = \nu_y \times \nu_x \times \nu_z$, *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 2.4 and 2.5 are modifications of Assumptions 2.1 and 2.2 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 2.3, 2.4 and 2.5. As in the previous section, properties of ν therefore translate into restrictions on $\mathbf{P}$. See the discussion preceding Theorem 2.1. The requirement $\mathbf{P} \subseteq \mathbf{M}_{y,x,z}(\nu)$ together with Assumption 2.3, rule out X_i being discrete under any $P \in \mathbf{P}$, such as in quantile treatment effect models with discrete treatments. See Lehmann (1975) and Chernozhukov and Hansen (2005). Finally, note that in each of the following two theorems Assumption 2.3 may also be relaxed in the same way as described preceding Theorem 2.1.

Theorem 2.2. *Suppose ν satisfies Assumptions 2.3, 2.4 and 2.5. Define $\mathbf{M}_{y,x,z}(\nu)$ as in (11) and let $\mathbf{P}$ be the maximal subset of $\mathbf{M}_{y,x,z}(\nu)$ such that for each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P) = L^\infty(P)$ for which (5) holds. Further define $\mathbf{P}_0$ and $\mathbf{P}_1$ as in (6). If a sequence of test functions $\{\phi_n\}_{n=1}^\infty$ satisfies (2) for some $\alpha \in (0,1)$, then it also satisfies (3).*

In establishing Theorem 2.2, 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 $H(P, P_k) = o(1)$. 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 Hellinger metric. As a result, the conclusion of Theorem 2.2 continues to hold if we instead set $\Theta(P) = L^q(P)$ for any $1 \leq q < \infty$. It is worth noting that, in contrast to the setting of Theorem 2.1, here letting $\Theta(P) = L^q(P)$ for $1 \leq q < \infty$ enlarges $\mathbf{P}$ itself, and so potentially enlarges not only $\mathbf{P}_0$, but also $\mathbf{P}_1$.

Remark 2.4. In establishing the denseness of $\mathbf{P}_0$ in $\mathbf{P}_1$, 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 $L^\infty(P_k)$ 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. ∎

Theorem 2.3. *Suppose ν satisfies Assumptions 2.3, 2.4 and 2.5. Define $\mathbf{M}_{y,x,z}(\nu)$ as in (11) and let $\mathbf{P}$ be the maximal subset of $\mathbf{M}_{y,x,z}(\nu)$ such that for each $P \in \mathbf{P}$ there is some $\theta \in \Theta(P) = \mathbf{T}(P)$ for which (7) holds, where $\mathbf{T}(P)$ is defined in (12). Further define $\mathbf{P}_0$ and $\mathbf{P}_1$ as in (8). If a sequence of test functions $\{\phi_n\}_{n=1}^\infty$ satisfies (2) for some $\alpha \in (0,1)$, then it also satisfies (3).*

As in Theorem 2.2, our proof implies that $\mathbf{P}_0$ is dense in $\mathbf{M}_{y,x,z}(\nu)$ with respect to the Hellinger metric. It therefore follows that the conclusion of Theorem 2.3 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 $1 \le q \le \infty$. 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 $\mathbf{T}(P_k)$ satisfying

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

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 2.4.

3   Conclusion

This paper has provided conditions under which nontrivial tests do not exist for each of three distinct hypothesis testing problems that arise in the context of identification of some nonparametric models with endogeneity. The first hypothesis testing problem considered concerns 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 consider concern testing identification directly in some nonparametric models with endogeneity involving quantile independence restrictions. Importantly, our conditions are satisfied under commonly used assumptions. On the other hand, they do not rule out the existence of reasonable tests under more restrictive assumptions. For instance, our arguments may not extend easily to cases where θ is known to satisfy additional shape restrictions or θ lies in a (pre-specified) compact subset of a suitable function space. In this way, our results may help shape the development of nontrivial tests of the hypotheses we consider.

4 Appendix

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

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

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

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

$L^q(\lambda)$ For $1 \leq q \leq \infty$ and a measure λ, the space $L^q(\lambda) \equiv \{f : \|f\|_{L^q(\lambda)} < \infty\}$.

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

PROOF: We proceed by constructing a map $B : [0,1] \to \mathcal{A}$ on a dense subset of $[0,1]$ and extending it to the entire domain $[0,1]$. Towards this end, let $\mathcal{F}_n \equiv \{0, \frac{1}{2^n}, \frac{2}{2^n}, \frac{3}{2^n}, \ldots, \frac{2^n-1}{2^n}, 1\}$, denote $\mathcal{F} \equiv \bigcup_{n=1}^\infty \mathcal{F}_n$ and define a map $\Gamma : \mathcal{A} \to \mathcal{A}$ that assigns to each $C \in \mathcal{A}$ a set $\Gamma(C) \subseteq C$ satisfying:

$$\lambda\{\Gamma(C)\} = \frac{1}{2}\lambda\{C\} , \tag{13}$$

where the existence of such $\Gamma(C)$ is ensured by Corollary 1.12.10 in Bogachev (2007) and λ being atomless. On $\mathcal{F}_1 = \{0, \frac{1}{2}, 1\}$, then define a map $B_1 : \mathcal{F}_1 \to \mathcal{A}$ by setting $B_1(0) = \emptyset$, $B_1(1) = A$ and $B_1(\frac{1}{2}) = \Gamma(A)$. Proceeding inductively, we then construct $B_n : \mathcal{F}_n \to \mathcal{A}$ by letting:

$$B_n\left(\frac{j}{2^n}\right) = \begin{cases} B_{n-1}\left(\frac{j/2}{2^{n-1}}\right) & \text{if } 1 \leq j \leq 2^n \text{ is even} \\ B_{n-1}\left(\frac{(j-1)/2}{2^{n-1}}\right) \cup \Gamma\left(B_{n-1}\left(\frac{(j+1)/2}{2^{n-1}}\right) \setminus B_{n-1}\left(\frac{(j-1)/2}{2^{n-1}}\right)\right) & \text{if } 1 \leq j \leq 2^n \text{ is odd} \end{cases} , \tag{14}$$

where in (14) we have exploited that $\tau \in \mathcal{F}_n$ if and only if $\tau = \frac{j}{2^n}$ for some integer $1 \leq j \leq 2^n$. Notice that if $\tau \in \mathcal{F}_n$, then it is of the form $\frac{j}{2^n} = \frac{2^m j}{2^{n+m}} \in \mathcal{F}_{n+m}$ for any integer $m \geq 0$ and hence:

$$B_{n+m}(\tau) = B_n(\tau) \;\; \text{for all } \tau \in \mathcal{F}_n , \tag{15}$$

as a result of definition (14). We may then define a map $B : \mathcal{F} \to \mathcal{A}$, pointwise given by:

$$B(\tau) = B_n(\tau) , \tag{16}$$

for any n such that $\tau \in \mathcal{F}_n$, and note $B(\tau)$ is uniquely determined due to (15).

Next, observe that since $B_1(0) = \emptyset$ and $B_1(1) = A$, it follows from (16) that $B(0) = \emptyset$ and $B(1) = A$ as well. Moreover, by induction it is also possible to establish that for all n:

$$\lambda\{B_n(\tau)\} = \tau\lambda\{A\} \quad \text{for all } \tau \in \mathcal{F}_n . \tag{17}$$

To see this, note (17) trivially holds for $n = 1$. Supposing (17) also holds for $n - 1$, then (14) verifies it must be satisfied for n and all $\tau \in \mathcal{F}_n$ of the form $\tau = \frac{j}{2^n}$ with $1 \le j \le 2^n$ even. On the other hand, if $\tau \in \mathcal{F}_n$ satisfies $\tau = \frac{j}{2^n}$ for $1 \le j \le 2^n$ odd, then by (14) we obtain,

$$
\begin{aligned}
\lambda\{B_n(\frac{j}{2^n})\} &= \lambda\{B_{n-1}(\frac{(j-1)}{2^n})\} + \lambda\{\Gamma(B_{n-1}(\frac{j+1}{2^n}) \setminus B_{n-1}(\frac{j-1}{2^n}))\} \\
&= \frac{(j-1)}{2^n}\lambda\{A\} + \frac{1}{2}(\frac{(j+1)}{2^n}\lambda\{A\} - \frac{(j-1)}{2^n}\lambda\{A\}) = \frac{j}{2^n}\lambda\{A\} , \quad (18)
\end{aligned}
$$

where the second equality follows by (13) and the induction hypothesis. We conclude (17) holds, and thus by (16) that $\lambda\{B(\tau)\} = \tau\lambda\{A\}$ for all $\tau \in \mathcal{F}$. Similarly, we may show inductively

$$
B_n(\tau) \subseteq B_n(\tau') \quad \text{for all } \tau \le \tau', \ \tau, \tau' \in \mathcal{F}_n . \tag{19}
$$

Property (19) is trivially satisfied by $B_1 : \mathcal{F}_1 \to \mathcal{A}$. To establish it holds for n, let $\tau' = \frac{j}{2^n}$ and $\tau = \frac{j-1}{2^n}$ for any integer $1 \le j \le 2^n$. If j is odd, then by definition (14) we may conclude:

$$
B_n(\tau') \supseteq B_{n-1}(\frac{(j-1)/2}{2^{n-1}}) = B_n(\frac{j-1}{2^n}) = B_n(\tau) . \tag{20}
$$

On the other hand, if j is even, then by (14) and the induction hypothesis, we can obtain:

$$
B_n(\tau) = B_{n-1}(\frac{j-2}{2^n}) \cup \Gamma(B_{n-1}(\frac{j}{2^n}) \setminus B_{n-1}(\frac{j-2}{2^n})) \subseteq B_{n-1}(\frac{j/2}{2^{n-1}}) = B_n(\tau') , \tag{21}
$$

where to derive the inclusion we have used that $\Gamma(C) \subseteq C$ for every $C \in \mathcal{A}$. Thus, since $1 \le j \le 2^n$ was arbitrary, from (20) and (21), we conclude that (19) holds. By construction, it additionally follows from (16) that $B : \mathcal{F} \to \mathcal{A}$ satisfies $B(\tau) \subseteq B(\tau')$ for all $\tau \le \tau'$ with $\tau, \tau' \in \mathcal{F}$.

To establish the first claim, we obtain $\bar{B} : [0,1] \to \mathcal{A}$ by extending $B : \mathcal{F} \to \mathcal{A}$. Specifically, for any $\tau \in \mathcal{F}$, let $\bar{B}(\tau) = B(\tau)$. For any $\tau \in [0,1] \setminus \mathcal{F}$ note that $0 \in \mathcal{F}$ and $\mathcal{F}$ being dense in $[0,1]$ imply we may select a sequence $\{\tau_j\}_{j=1}^\infty$ with $\tau_j \in \mathcal{F}$ for all j, such that $\tau_j \uparrow \tau$. Then define:

$$
\bar{B}(\tau) = \bigcup_{j=1}^\infty B(\tau_j) . \tag{22}
$$

Notice that the definition of $\bar{B}(\tau)$ is independent of the sequence $\{\tau_j\}_{j=1}^\infty$ due to $B(\tau) \subseteq B(\tau')$ for any $\tau \le \tau'$ with $\tau, \tau' \in \mathcal{F}$, and $\{\tau_j\}_{j=1}^\infty$ approaching τ from below. Also note that since $\{0, 1\} \in \mathcal{F}$, $\bar{B}(0) = B(0) = \emptyset$ and $\bar{B}(1) = B(1) = A$. In addition, by the monotone convergence theorem:

$$
\lambda\{\bar{B}(\tau)\} = \lambda\{\bigcup_{j=1}^\infty B(\tau_j)\} = \lim_{j\to\infty} \lambda\{B(\tau_j)\} = \lim_{j\to\infty} \tau_j \lambda\{A\} = \tau\lambda\{A\} , \tag{23}
$$

where in the third equality we have exploited $\lambda\{B(\tau)\} = \tau\lambda\{A\}$ for all $\tau \in \mathcal{F}$. Finally, since $\mathcal{F}$ is dense in $[0,1]$, we have that for any $0 < \tau < \tau'$, there exist sequences $\{\tau_j\}_{j=1}^\infty$ and $\{\tau_j'\}_{j=1}^\infty$ with $\tau_j \uparrow \tau$, $\tau_j' \uparrow \tau'$ and $\tau_j, \tau_j' \in \mathcal{F}$ for all j. Selecting a $\bar{\tau} \in \mathcal{F}$ such that $\tau < \bar{\tau} < \tau'$ we then obtain:

$$
\bar{B}(\tau) = \bigcup_{j=1}^\infty B(\tau_j) \subseteq B(\bar{\tau}) \subseteq \bigcup_{j=1}^\infty B(\tau_j') = \bar{B}(\tau') , \tag{24}
$$

where we have used that $\tau_j \uparrow \tau < \bar{\tau}$ implies $B(\tau_j) \subseteq B(\bar{\tau})$ for all j, and similarly, that $\tau'_j \uparrow \tau' > \bar{\tau}$ implies $B(\bar{\tau}) \subseteq B(\tau'_j)$ for j sufficiently large. We conclude from (23) and (24) that $\bar{B} : [0,1] \to \mathcal{A}$ additionally satisfies properties (ii) and (iii), and the first claim of the Lemma is established.

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

$$\tilde{B}(\tau) = A \setminus \bar{B}(1 - \tau) . \tag{25}$$

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

$$\lambda\{\tilde{B}(\tau)\} = \lambda\{A \setminus \bar{B}(1 - \tau)\} = \lambda\{A\} - (1 - \tau)\lambda\{A\} = \tau\lambda\{A\} . \tag{26}$$

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:

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

Thus, from (26) and (27) we obtain that $\tilde{B} : [0,1] \to \mathcal{A}$ 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:

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

Therefore, it follows from (28) that if $\lambda\{A\} > 0$, then $\lambda\{\tilde{B}(\tau) \triangle \bar{B}(\tau)\} > 0$ for all $\tau \in (0,1)$. $\blacksquare$

Lemma A.2. *Suppose $Q \in \mathbf{M}$ satisfies $Q \ll \lambda$ for λ a σ-finite positive Borel measure on $\mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ and let $f \equiv \sqrt{dQ/d\lambda}$. Then, there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of simple functions:*

$$f_n(y, x, z) = \sum_{i=1}^{K_n} \pi_{in} 1\{(y, x, z) \in S_{in}\} ,$$

such that $\|f_n - f\|_{L^2(\lambda)} = o(1)$. Additionally, for all n: (i) $f_n \ge 0$ and $\int f_n^2 d\lambda = 1$, (ii) $\{S_{in}\}_{i=1}^{K_n}$ is a partition of $[-M_n, M_n]^{1+d_x+d_z}$ for some $M_n > 0$, (iii) For all $1 \le i \le K_n$, $S_{in} = A_{in} \times B_{in} \times C_{in}$ for some $A_{in} \subseteq [-M_n, M_n]$, $B_{in} \subseteq [-M_n, M_n]^{d_x}$ and $C_{in} \subseteq [-M_n, M_n]^{d_z}$, (iv) The distinct elements of $\{A_{in}\}_{i=1}^{K_n}$, $\{B_{in}\}_{i=1}^{K_n}$ and $\{C_{in}\}_{i=1}^{K_n}$ form a partition of $[-M_n, M_n]$, $[-M_n, M_n]^{d_x}$ and $[-M_n, M_n]^{d_z}$.

PROOF: Note that by construction $f \in L^2(\lambda)$. Since λ is a Borel regular measure by assumption, Theorem 13.9 in Aliprantis and Border (2006) implies there exists a sequence $\{\bar{f}_n\}_{n=1}^{\infty}$ of continuous, compactly supported functions such that $\|\bar{f}_n - f\|_{L^2(\lambda)} = o(1)$. Moreover, since $f \ge 0$ we may assume without loss of generality that $\bar{f}_n \ge 0$ for all n, while $\|f\|_{L^2(\lambda)} = 1$ and $\|\bar{f}_n - f\|_{L^2(\lambda)} = o(1)$ further imply that $\|\bar{f}_n\|_{L^2(\lambda)} \to 1$. Therefore, defining $\tilde{f}_n \equiv \bar{f}_n / \|\bar{f}_n\|_{L^2(\lambda)}$ we obtain that:

$$\|\tilde{f}_n - f\|_{L^2(\lambda)} \le \frac{\|\bar{f}_n - f\|_{L^2(\lambda)}}{\|\bar{f}_n\|_{L^2(\lambda)}} + |1 - \frac{1}{\|\bar{f}_n\|_{L^2(\lambda)}}| \times \|f\|_{L^2(\lambda)} = o(1) . \tag{29}$$

Let $\Omega_n \subset \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$ be the support of $\tilde{f}_n$, which is compact due to $\bar{f}_n$ being compactly supported, and select $M_n > 0$ sufficiently large so that $\Omega_n \subseteq [-M_n, M_n]^{1+d_x+d_z}$. Additionally, select

11

$\xi_n \downarrow 0$ such that $\lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(\xi_n^{-2})$, and notice that $\tilde{f}_n$ being uniformly continuous on $[-M_n, M_n]^{1+d_x+d_z}$ implies there exists a partition $\{S_{in}\}_{i=1}^{K_n}$ of $[-M_n, M_n]^{1+d_x+d_z}$ such that:

$$\max_{1 \leq i \leq K_n} \sup_{(y,x,z) \in S_{in}} \sup_{(y',x',z') \in S_{in}} |\tilde{f}_n(y, x, z) - \tilde{f}_n(y', x', z')| \leq \xi_n . \tag{30}$$

Moreover, without loss of generality, we may construct $\{S_{in}\}_{i=1}^{K_n}$ through the product of partitions of $[-M_n, M_n]$, $[-M_n, M_n]^{d_x}$ and $[-M_n, M_n]^{d_z}$ so that conditions (iii)-(iv) are satisfied. Next, let

$$\pi_{in} = \begin{cases} [(\int_{S_{in}} \tilde{f}_n^2 d\lambda)/\lambda\{S_{in}\}]^{\frac{1}{2}} & \text{if } \lambda\{S_{in}\} \neq 0 \\ 0 & \text{if } \lambda\{S_{in}\} = 0 \end{cases} . \tag{31}$$

We may then obtain the desired functions $f_n : \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z} \to \mathbf{R}$, defining them pointwise by

$$f_n(y, x, z) = \sum_{i=1}^{K_n} \pi_{in} 1\{(y, x, z) \in S_{in}\} . \tag{32}$$

By construction, $\pi_{in} \geq 0$ for all $1 \leq i \leq K_n$ and all n, and hence $f_n \geq 0$ as desired. Additionally, because the support of $\tilde{f}_n$ is contained in $[-M_n, M_n]^{1+d_x+d_z} = \bigcup_{i=1}^{K_n} S_{in}$ we also obtain by (31):

$$\int f_n^2 d\lambda = \sum_{i=1}^{K_n} \pi_{in}^2 \lambda\{S_{in}\} = \sum_{i=1}^{K_n} \int_{S_{in}} \tilde{f}_n^2 d\lambda = \int \tilde{f}_n^2 d\lambda = 1 . \tag{33}$$

Furthermore, for all n and each $1 \leq i \leq K_n$ such that $\lambda\{S_{in}\} > 0$, we obtain from (30) and (31):

$$\sup_{(y,x,z) \in S_{in}} |f_n(y, x, z) - \tilde{f}_n(y, x, z)| = \sup_{(y,x,z) \in S_{in}} |[\frac{\int_{S_{in}} \tilde{f}_n^2 d\lambda}{\lambda\{S_{in}\}}]^{\frac{1}{2}} - \tilde{f}_n(y, x, z)| \leq \xi_n . \tag{34}$$

Hence, since $f_n, \tilde{f}_n$ vanish outside $[-M_n, M_n]^{1+d_x+d_z}$, and $\{S_{in}\}_{i=1}^{K_n}$ is a partition of $[-M_n, M_n]^{1+d_x+d_z}$ we may conclude from (34) that $\sup_{(y,x,z)} |f_n(y, x, z) - \tilde{f}_n(y, x, z)| \leq \xi_n$ λ-a.s. Thus, we obtain:

$$\int (f_n - \tilde{f}_n)^2 d\lambda = \int_{[-M_n, M_n]^{1+d_x+d_z}} (f_n - \tilde{f}_n)^2 d\lambda \leq \xi_n^2 \lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(1) , \tag{35}$$

where the final equality holds because we selected $\xi_n \downarrow 0$ such that $\lambda\{[-M_n, M_n]^{1+d_x+d_z}\} = o(\xi_n^{-2})$. Combining (29) and (35) then implies $\|f_n - f\|_{L^2(\lambda)} = o(1)$ establishing the claim of the Lemma. ∎

PROOF OF LEMMA 2.1: Fix $P \in \mathbf{P}_1$ and let $\{P_k\}_{k=1}^{\infty}$ satisfy $H(P_k, P) = o(1)$ and $P_k \in \mathbf{P}_0$ for all k. For every fixed n, problem 4.18 in Pollard (2006) then implies that:

$$\lim_{k \to \infty} H(P_k^n, P^n) \leq \lim_{k \to \infty} nH(P_k, P) = 0 . \tag{36}$$

Since $P \in \mathbf{P}_1$ was arbitrary, we conclude from (36) that for every n and $P \in \mathbf{P}_1$ there exists a sequence $\{P_k\}_{k=1}^{\infty}$ in $\mathbf{P}_0$ such that $H(P_k^n, P^n) = o(1)$. Moreover, by Theorem 4.2.37 in Bogachev (2007), the Hellinger distance induces the same topology on $\mathbf{M}$ as the Total Variation metric, and

12

therefore it follows by Theorem 1 in Romano (2004) that for every n and bounded function ϕ_n we have:

$$\sup_{P \in \mathbf{P}_1} E_{P^n}[\phi_n] \leq \sup_{P \in \mathbf{P}_0} E_{P^n}[\phi_n] . \tag{37}$$

The conclusion of the Lemma is then immediate from (37) holding for all n. ∎

PROOF OF THEOREM 2.1: Fix $P \in \mathbf{P}_1$ and let $f \equiv dP/d\nu$. By Assumption 2.1 and Lemma A.2 applied to $\lambda = \delta \times \nu$ for δ a degenerate measure on $\mathbf{R}$, there exists a sequence $\{f_k\}_{k=1}^{\infty}$ such that $\|\sqrt{f} - \sqrt{f_k}\|_{L^2(\nu)} = o(1)$, and each f_k is a simple function of the form:

$$f_k(x, z) = \sum_{i=1}^{K_k} \pi_{ik}^2 1\{(x, z) \in S_{ik}\} , \tag{38}$$

with $\{S_{ik}\}_{i=1}^{K_k}$ a partition of $[-M_k, M_k]^{d_x+d_z}$ for some $M_k > 0$ and $S_{ik} = B_{ik} \times C_{ik}$ for some $B_{ik} \subseteq [-M_k, M_k]^{d_x}$, $C_{ik} \subseteq [-M_k, M_k]^{d_z}$ for all k and $1 \leq i \leq K_k$. Also define,

$$P_k\{E\} \equiv \int_E f_k d\nu \tag{39}$$

for all Borel measurable $E \subseteq \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, and note that since $f_k \geq 0$ and $\int f_k d\nu = 1$ by Lemma A.2, while $P_k \ll \nu$ by construction, it follows that $P_k \in \mathbf{P} = \mathbf{M}_{x,z}(\nu)$ for all k. Moreover, since ν is a common dominating measure for P and P_k for all k, we obtain $H(P, P_k) = \|\sqrt{f_k} - \sqrt{f}\|_{L^2(\nu)} = o(1)$, and hence $P_k \to P$ with respect to the Hellinger metric.

In what follows, we aim to show that in fact $P_k \in \mathbf{P}_0$ for all k. Towards this end, let $\{U_{ik}\}_{i=1}^{D_k}$ denote the collection of distinct elements of $\{B_{ik}\}_{i=1}^{K_k}$. Assumption 2.3 and Corollary 1.12.10 in Bogachev (2007), then imply that for each U_{ik} there exist Borel measurable subsets $(U_{ik}^{(1)}, U_{ik}^{(2)})$ such that $U_{ik} = U_{ik}^{(1)} \cup U_{ik}^{(2)}$, $U_{ik}^{(1)} \cap U_{ik}^{(2)} = \emptyset$ and in addition satisfy:

$$\nu_x\{U_{ik}^{(1)}\} = \nu_x\{U_{ik}^{(2)}\} = \frac{1}{2}\nu_x\{U_{ik}\} . \tag{40}$$

Since $\{U_{ik}\}_{k=1}^{D_k}$ is a partition of $[-M_k, M_k]^{d_x}$ by Lemma A.2, we may define a function θ_k by:

$$\theta_k(x) = \sum_{i=1}^{D_k} (1\{x \in U_{ik}^{(1)}\} - 1\{x \in U_{ik}^{(2)}\}) , \tag{41}$$

and note that $\nu_x\{x \in [-M_k, M_k]^{d_x} : \theta_k(x) = 0\} = 0$ due to (40), and $U_{ik}^{(1)} \cap U_{ik}^{(2)} = \emptyset$ for all $1 \leq i \leq D_k$. Hence, $P_k \ll \nu$, and the support of X_i under P_k being contained in $[-M_k, M_k]^{d_x}$ implies $\theta_k \neq 0$ P_k-a.s., while θ_k being bounded yields $\theta_k \in L^{\infty}(P_k)$. Additionally, for any bounded $z \mapsto \psi(z)$, we obtain from (38), $f_k = dP_k/d\nu$ and Assumption 2.3 that:

$$E_{P_k}[\psi(Z_i)\theta_k(X_i)] = \sum_{i=1}^{K_k} \pi_{ik}^2 \int_{C_{ik}} \int_{B_{ik}} \psi(z)\theta_k(x)\nu_x(dx)\nu_z(dz)$$
$$= \sum_{i=1}^{K_k} \pi_{ik}^2 (\sum_{j=1}^{D_k} \nu_x\{B_{ik} \cap U_{jk}^{(1)}\} - \nu_x\{B_{ik} \cap U_{jk}^{(2)}\}) \int_{C_{ik}} \psi(z)\nu_z(dz) = 0 , \quad (42)$$

13

where for the final equality we have exploited (40) and that for every $1 \leq i \leq K_k$, we have $B_{ik} = U_{jk}$ for some $1 \leq j \leq D_k$. In particular, (42) 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.

Thus, we may conclude from (42) that $P_k \in \mathbf{P}_0$ for all k. Hence, since $P \in \mathbf{P}_1$ was arbitrary and $H(P, P_k) = o(1)$, the conclusion of the Theorem follows by Lemma 2.1 and (2). ∎

PROOF OF THEOREM 2.2: Fix $P \in \mathbf{P}_1$ and let $f \equiv dP/d\nu$. By Lemma A.2 and Assumption 2.4, there exists a sequence $\{f_k\}_{k=1}^\infty$ such that $\|\sqrt{f} - \sqrt{f_k}\|_{L^2(\nu)} = o(1)$ and each f_k is of the form:

$$f_k(y, x, z) = \sum_{i=1}^{K_k} \pi_{ik}^2 1\{(y, x, z) \in S_{ik}\} , \tag{43}$$

with $\{S_{ik}\}_{i=1}^{K_k}$ a partition of $[-M_k, M_k]^{1+d_x+d_z}$ for some $M_k > 0$ and $S_{ik} = A_{ik} \times B_{ik} \times C_{ik}$ for some $A_{ik} \subseteq [-M_k, M_k]$, $B_{ik} \subseteq [-M_k, M_k]^{d_x}$, $C_{ik} \subseteq [-M_k, M_k]^{d_z}$ for all k and $1 \leq i \leq K_k$. Also define,

$$P_k\{E\} \equiv \int_E f_k d\nu \tag{44}$$

for all Borel measurable $E \subseteq \mathbf{R} \times \mathbf{R}^{d_x} \times \mathbf{R}^{d_z}$, and note that since $f_k \geq 0$ and $\int f_k d\nu = 1$ by Lemma A.2, while $P_k \ll \nu$ by construction, it follows that $P_k \in \mathbf{M}_{y,x,z}(\nu)$ for all k. Moreover, since ν is a common dominating measure for P and P_k for all k, we obtain $H(P, P_k) = \|\sqrt{f_k} - \sqrt{f}\|_{L^2(\nu)} = o(1)$, and hence $P_k \to P$ with respect to the Hellinger metric.

Next, let $\{U_{ik}\}_{i=1}^{D_k}$ denote the collection of distinct elements of $\{B_{ik}\}_{i=1}^{K_k}$. Assumption 2.3 and Lemma A.1 then imply there exist collections $\{U_{ik}^{(1)}(\tau), U_{ik}^{(2)}(\tau)\}_{i=1}^{D_k}$ such that for all $1 \leq i \leq D_k$,

$$\nu_x\{U_{ik}^{(1)}(\tau)\} = \tau\nu_x\{U_{ik}\} \qquad \nu_x\{U_{ik}^{(2)}(\tau)\} = \tau\nu_x\{U_{ik}\} , \tag{45}$$

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

$$\theta_k^{(l)}(x, \tau) = \sum_{i=1}^{D_k} (2M_k 1\{x \in U_{ik}^{(l)}(\tau)\} - 2M_k 1\{x \in U_{ik} \setminus U_{ik}^{(l)}(\tau)\}) , \tag{46}$$

where we note that since $\{U_{ik}\}_{i=1}^{D_k}$ is a partition of $[-M_k, M_k]^{d_x}$ by Lemma A.2, it follows that $\theta_k^{(l)}(x, \tau) \in \{-2M_k, 2M_k\}$ for all $x \in [-M_k, M_k]^{d_x}$. Thus, $\theta_k^{(l)} \in L^\infty(P_k)$ for $l \in \{1, 2\}$, while $Y_i \in [-M_k, M_k]$ P_k-a.s., Assumption 2.5 and $f_k = dP_k/d\nu$ with f_k as in (43) additionally imply:

$$E_{P_k}[(1\{Y_i \leq \theta_k^{(1)}(X_i, \tau)\} - 1\{Y_i \leq \theta_k^{(2)}(X_i, \tau)\})^2]$$
$$= E_{P_k}[(1\{\theta_k^{(1)}(X_i, \tau) = 2M_k\} - 1\{\theta_k^{(2)}(X_i, \tau) = 2M_k\})^2]$$
$$= \sum_{i=1}^{K_k} \pi_{ik}^2 \nu_y\{A_{ik}\} \nu_z\{C_{ik}\} \int_{B_{ik}} (1\{\theta_k^{(1)}(x, \tau) = 2M_k\} - 1\{\theta_k^{(2)}(x, \tau) = 2M_k\})^2 \nu_x(dx)$$
$$> 0 , \tag{47}$$

14

where we exploited $(1\{\theta_k^{(1)}(x,\tau) = 2M_k\} - 1\{\theta_k^{(2)}(x,\tau) = 2M_k\})^2 = 1\{x \in U_{jk}^{(1)}(\tau) \triangle U_{jk}^{(2)}(\tau)\}$ for every $x \in U_{jk}$, $\nu_x\{U_{jk}^{(1)}(\tau) \triangle U_{jk}^{(2)}(\tau)\} > 0$ whenever $\nu_x\{U_{jk}\} > 0$, and $\nu_x\{U_{jk}\} > 0$ for some j due to the support of X_i under P_k being contained in $[-M_k, M_k]^{d_x} = \bigcup_{j=1}^{D_k} U_{jk}$.

We conclude from (47) 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 $1\{\theta_k^{(l)}(x,\tau) = 2M_k\} = 1\{x \in U_{jk}^{(l)}(\tau)\}$ for every $x \in U_{jk}$ by (46), and hence $\nu_x\{\{\theta_k^{(l)}(x,\tau) = 2M_k\} \cap U_{jk}\} = \nu_x\{U_{jk}^{(l)}(\tau)\} = \tau \nu_x\{U_{jk}\}$. Since for any $1 \le i \le K_k$, $B_{ik} = U_{jk}$ for some $1 \le j \le D_k$, and $[-M_k, M_k] = \bigcup_{i=1}^{K_k} A_{ik}$, it follows that for any bounded $z \mapsto \psi(z)$ we obtain:

$$
E_{P_k}[\psi(Z_i)(1\{Y_i \le \theta_k^{(l)}(X_i, \tau)\} - \tau)]
$$
$$
= \sum_{i=1}^{K_k} \pi_{ik}^2 \int_{A_{ik}} \int_{C_{ik}} \int_{B_{ik}} \psi(z)(1\{\theta_k^{(l)}(x,\tau) = 2M_k\} - \tau)\nu_x(dx)\nu_z(dz)\nu_y(dy)
$$
$$
= \sum_{i=1}^{K_k} \pi_{ik}^2 (\tau \nu_x\{B_{ik}\} - \tau \nu_x\{B_{ik}\}) \int_{A_{ik}} \int_{C_{ik}} \psi(z)\nu_z(dz)\nu_y(dy)
$$
$$
= 0 \ . \tag{48}
$$

In particular, setting $\psi(\cdot) = P_k\{Y_i \le \theta^{(l)}(X_i, \tau)|Z_i = \cdot\}$ in (48), implies by the law of iterated expectations that $P_k\{Y_i \le \theta_k^{(l)}(X_i, \tau)|Z\} = \tau$, $P_k - a.s.$ for $l \in \{1,2\}$.

Thus, we may conclude from (42) that $P_k \in \mathbf{P}_0$ for all k. Hence, since $P \in \mathbf{P}_1$ was arbitrary and $H(P, P_k) = o(1)$, the conclusion of the Theorem follows by Lemma 2.1 and (2). ∎

PROOF OF THEOREM 2.3: The proof is very similar to that of Theorem 2.2, and we therefore provide only an outline, emphasizing the differences in the arguments. Fixing $\mathbf{P} \in \mathbf{P}_1$, 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 (43), and $\|\sqrt{f_k} - \sqrt{f}\|_{L^2(\nu)} = o(1)$. To show $P_k \in \mathbf{P}_0$ for all k, let $\{U_{ik}\}_{i=1}^{D_k}$ denote the collection of unique elements of $\{B_{ik}\}_{i=1}^{K_k}$, and for all $1 \le i \le D_k$, let $\mathcal{U}_{ik}$ denote the σ-algebra generated by subsets of U_{ik}. By Assumption 2.3 and Lemma A.1, there then exist $U_{ik}^{(1)} : [0,1] \to \mathcal{U}_{ik}$ and $U_{ik}^{(2)} : [0,1] \to \mathcal{U}_{ik}$ such that for $l \in \{1,2\}$: (i) $\nu_x\{U_{ik}^{(l)}(\tau)\} = \tau \nu_x\{U_{ik}\}$, (ii) $U_{ik}^{(l)}(\tau) \subseteq U_{ik}^{(l)}(\tau')$ for all $0 \le \tau \le \tau' \le 1$, and (iii) $\nu_x\{U_{ik}^{(1)}(\tau) \triangle U_{ik}^{(2)}(\tau)\} > 0$ for all $\tau \in (0,1)$ and $1 \le i \le D_k$ such that $\nu_x\{U_{ik}\} > 0$. Following (46), we can then define the functions $\theta_k^{(l)}$ pointwise by:

$$
\theta_k^{(l)}(x, \tau) = \sum_{i=1}^{D_k} ((2+\tau)M_k 1\{x \in U_{ik}^{(l)}(\tau)\} - (3-\tau)M_k 1\{x \in U_{ik} \setminus U_{ik}^{(l)}(\tau)\}) \ . \tag{49}
$$

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 $U_{ik}^{(l)}(\tau) \subseteq U_{ik}^{(l)}(\tau')$ for $l \in \{1,2\}$ and all $0 \le \tau \le \tau' \le 1$ and $1 \le i \le D_k$, it follows from (49) and the support of X_i under P_k being contained in $[-M_k, M_k]^{d_x} = \bigcup_{j=1}^{D_k} U_{jk}$ 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 (49) that for all $x \in U_{ik}^{(l)}$, $1\{Y_i \le \theta_k^{(l)}(x, \tau)\} = 1\{x \in U_{ik}^{(l)}(\tau)\}$ P_k-a.s. Therefore, arguing as in (47) yields that:

$$
E_{P_k}[(1\{Y_i \le \theta_k^{(1)}(X_i, \tau)\} - 1\{Y_i \le \theta_k^{(2)}(X_i, \tau)\})^2] > 0 \ , \tag{50}
$$

for all $\tau \in (0, 1)$. Thus, we may conclude from (50) that for all $\tau \in (0, 1)$ $\theta_k^{(1)}(\cdot, \tau)$ differs from $\theta_k^{(2)}(\cdot, \tau)$ under $\| \cdot \|_{L^\infty(P_k)}$. Similarly, arguing as in (48) further implies $P_k\{Y_i \le \theta^{(l)}(X_i, \tau)|Z_i\} = \tau$ for $l \in \{1, 2\}$ and all $\tau \in (0, 1)$ P_k-a.s. Therefore, we may conclude $P_k \in \mathbf{P}_0$ for all k, and finish the argument as in the proof of Theorem 2.2. ∎

References

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

ANDERSON, T. (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.

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

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

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

CHEN, X. and POUZO, D. (2008). Estimation of nonparametric conditional moment models with possibly nonsmooth moments.

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.

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. (2010). 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.

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.

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

POLLARD, D. (2006). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.

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. (2010). Inference in nonparametric instrumental variables with partial identification. Working paper, University of California - San Diego.

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