On completeness and consistency in nonparametric instrumental variable models*

Joachim Freyberger[‡]

April 14, 2017

Abstract

This paper provides positive testability results for the identification condition in a nonparametric instrumental variable model, known as completeness, and it links the outcome of the test to properties of an estimator of the structural function. In particular, I show that the data can provide empirical evidence in favor of both an arbitrarily small identified set as well as an arbitrarily small asymptotic bias of the estimator. This is the case for a large class of complete distributions as well as certain incomplete distributions. As a byproduct, the results can be used to estimate an upper bound of the diameter of the identified set and to obtain an easy to report estimator of the identified set itself.

Keywords: Completeness, testing, consistency, instrumental variables, nonparametric estimation.

JEL Classification: C14, C21

*I thank Alberto Abadie and three anonymous referees for valuable suggestions which helped to substantially improve the paper. I also thank Ivan Canay, Bruce Hansen, Joel Horowitz, Jack Porter, Azeem Shaikh, Xiaoxia Shi, and seminar participants at various conferences and seminars for helpful comments and discussions. Jangsu Yoon provided excellent research assistance.

[‡]Department of Economics, University of Wisconsin - Madison. Email: jfreyberger@ssc.wisc.edu.

1 Introduction

There has been much recent work on nonparametric models with endogeneity, which relies on a nonparametric analog of the rank condition, known as completeness. Specifically, consider the nonparametric instrumental variable (NPIV) model

$$Y = g_0(X) + U, \qquad E(U \mid Z) = 0, \tag{1}$$

where Y, X, and Z are observed scalar random variables, U is an unobserved random variable, and g_0 is a structural function of interest. It is well known that identification in this model is equivalent to the completeness condition (Newey and Powell, 2003), which says that $E(g(X) \mid Z) = 0$ almost surely implies that $g(X) = 0$ almost surely for all g in a certain class of functions.[1] Next to this NPIV model, completeness has also been used in various other settings including measurement error models (Hu and Schennach, 2008), panel data models (Freyberger, 2012), and nonadditive models with endogeneity (Chen, Chernozhukov, Lee, and Newey, 2014). Although completeness has been employed extensively, existing results so far have only established that the null hypothesis that completeness fails is not testable. In particular, Canay, Santos, and Shaikh (2013) show that any test that controls size uniformly over a large class of incomplete distributions, has power no greater than size against any alternative. Intuitively, the null hypothesis that completeness fails cannot be tested because for every complete distribution, there exists an incomplete distribution which is arbitrarily close to it. They conclude that "it is therefore not possible to provide empirical evidence in favor of the completeness condition by means of such a test".

In an application researchers most likely do not just want to test completeness by itself, but are instead interested in estimating g_0. One might expect that if an incomplete distribution is arbitrarily close to a complete distribution, a nonparametric estimator of g_0 has similar properties under both distributions. In particular, it turns out that even if completeness fails, it might be the case that the diameter of the identified set, denoted by $\text{diam}(I_0(P))$, is smaller than a fixed $\varepsilon > 0$.[2] It then follows

[1] The class of functions typically depends on the restrictions imposed on g_0, such as being square integrable ("L^2 completeness") or bounded ("bounded completeness").

[2] See Section 2 for a formal definition of the diameter of the identified set. To achieve a bounded identified set, the function g_0 has to satisfy commonly assumed smoothness restrictions; see Section 2.1 for details.

that for certain estimators $\hat{g}$ it holds that $\|\hat{g} - g_0\|_c \leq \varepsilon + o_p(1)$, where $\|\cdot\|_c$ is a consistency norm. In other words, for certain incomplete distributions, namely those close to complete distributions, $\hat{g}$ will be close to g_0 asymptotically.

In this paper I first show that under certain assumptions $H_0 : \text{diam}(I_0(P)) \geq \varepsilon$ is a testable hypothesis and that rejecting H_0 provides evidence in favor of a small asymptotic bias of a large class of estimators. Next, I formally link the outcome of a test to properties of an estimator. That is, I provide a test statistic $\hat{T}$, a critical value c_n, and an estimator $\hat{g}$, such that uniformly over a large class of distributions

$$P\left(\|\hat{g} - g_0\|_c \geq \varepsilon,\, n\hat{T} \geq c_n\right) \to 0$$

as $n \to \infty$, where n is the sample size. This result holds both for a fixed ε and when $\varepsilon \to 0$ as $n \to \infty$. Moreover, I show that $P(n\hat{T} \geq c_n) \to 1$ for a large class of complete distributions and certain sequences of incomplete distributions. An important implication of these results is that for any sequence of distributions for which $P(n\hat{T} \geq c_n) \geq \delta > 0$,

$$P\left(\|\hat{g} - g_0\|_c \geq \varepsilon \mid n\hat{T} \geq c_n\right) \to 0.$$

Hence, rejecting can provide evidence in favor of an arbitrarily small asymptotic bias of the estimator. Since the test does not control size uniformly over all incomplete distributions, the results imply that for certain sequences of incomplete distributions $\|\hat{g} - g_0\|_c \xrightarrow{p} 0$. Finally, I show how the test can be used to estimate an upper bound of the diameter of the identified set and to obtain an easy to report estimator of the identified set itself.

This paper does not address the question of conducting inference. Santos (2012) and Tao (2016) provide pointwise valid inference methods, which are robust to a failure of point identification, but they do not discuss properties of estimators of g_0 under partial identification, and they do not show that the data can provide evidence in favor of an arbitrarily small asymptotic bias or an arbitrarily small diameter of the identified set (see Section 4 for further discussion).

Literature: Most theoretical work in the NPIV literature relies on the completeness assumption, such as Newey and Powell (2003), Hall and Horowitz (2005), Blundell et al. (2007), Darolles et al. (2011), Horowitz (2011), Horowitz and Lee (2012), Horowitz (2014), and Chen and Christensen (2017). Other settings and applications which use completeness include Hu and Schennach (2008), Berry and Haile (2014), Chen et al. (2014), and Sasaki (2015). There is also a growing literature on general models with conditional moment restrictions, which include instrumental variable models as special cases. Several settings assume point identification (for example Ai and Chen (2003), Chen and Pouzo (2009, 2012, 2015)) while others allow for partial identification (Tao, 2016). Finally, there are several recent papers (including Mattner (1993), Newey and Powell (2003), Andrews (2011), D'Haultfoeuille (2011), and Hu and Shiu (2016)) which have provided sufficient conditions for different versions of completeness, such as bounded completeness or L^2 completeness. The results of Canay et al. (2013) imply that these versions of completeness are not testable, while the sufficient conditions might be testable if they are strong enough.

Section 2 provides definitions and a derivation of the population test statistic. Section 3 presents the sample analog and the main results which, among others, link the outcome of the test to properties of an estimator. All proofs are in the appendix. Additional material is in a supplementary appendix with section numbers S.1, etc..

2 Definitions and population test statistic

This section starts by introducing function spaces and norms that are used throughout the paper. It then explains the link between the diameter of the identified set and properties of estimators and it derives the population test statistic.

2.1 Notation

Let $\|\cdot\|$ be the Euclidean norm and let $\|\cdot\|_2$ denote the L^2-norm. Additionally, let $\mathcal{X}$ be the support of X and let $\|\cdot\|_c$ and $\|\cdot\|_s$ be two norms for functions from $\mathcal{X}$ to $\mathbb{R}$. Define the parameter space $\mathcal{G} = \{g : \|g\|_s \leq C\}$, where C is a positive constant. Properties of $\|\cdot\|_c$ and $\|\cdot\|_s$ are discussed below but useful examples to think of are:

$$\|g\|_c^2 = \int_{\mathcal{X}} g(x)^2 dx \quad \text{and} \quad \|g\|_s^2 = \int_{\mathcal{X}} \left(g(x)^2 + g'(x)^2\right) dx$$

or

$$\|g\|_c = \sup_{x\in\mathcal{X}} |g(x)| \quad \text{and} \quad \|g\|_s = \sup_{x\in\mathcal{X}} |g(x)| + \sup_{x_1,x_2\in\mathcal{X} x_1\neq x_2} |g(x_1) - g(x_2)|/\|x_1 - x_2\|.$$

A standard smoothness assumption in many nonparametric models, which I also impose in this paper, is that $g_0 \in \mathcal{G}$ (see e.g. Newey and Powell (2003), Santos (2012), or Horowitz (2014)). This assumption typically restricts function values and derivatives of g_0. Section S.3.2 in the supplement explains how these norm bounds can be derived in particular examples. Consistency is then usually proved in the weaker norm $\|\cdot\|_c$. It will also be convenient to define $\bar{\mathcal{G}}(\varepsilon) = \{g : \|g\|_s \leq 2C/\varepsilon\}$. With these restrictions define the identified set $I_0(P) = \{g \in \mathcal{G} : E(g(X) \mid Z) = E(Y \mid Z)\}$ and its diameter

$$\text{diam}(I_0(P)) = \sup_{g_1, g_2 \in I_0(P)} \|g_1 - g_2\|_c.$$[3]

2.2 Derivation of the population test statistic

I first show that if $\text{diam}(I_0(P)) \leq \varepsilon$, then the asymptotic bias of a large class of estimators will be small.[4] Specifically, let $\tilde{g}$ be any estimator such that

$$\inf_{g \in I_0(P)} \|\tilde{g} - g\|_c = o_p(1).$$

That is, $\tilde{g}$ is close to some function in the identified set as the sample size increases. Many estimators, such as series or Tikhonov estimators satisfy this property, even if g_0 is not point identified. Then if $\text{diam}(I_0(P)) \leq \varepsilon$,

$$\begin{aligned} \|\tilde{g} - g_0\|_c &= \inf_{g \in I_0(P)} \|\tilde{g} - g + g - g_0\|_c \\ &\leq \inf_{g \in I_0(P)} \|\tilde{g} - g\|_c + \sup_{g \in I_0(P)} \|g - g_0\|_c \\ &\leq o_p(1) + \varepsilon. \end{aligned}$$

For a fixed distribution of the data an estimator of g_0 is typically not consistent if g_0 is not point identified, but these derivations show that the asymptotic bias can be arbitrarily small. Moreover, for a sequence of distributions, $\hat{g}$ is consistent as long as $\text{diam}(I_0(P)) \to 0$ as $n \to \infty$.

[3]For some quantities, such as $I_0(P)$, I make the dependence on the distribution of the data P explicit, which will be important in Section 3.

[4]Since the asymptotic bias is guaranteed to be small, this situations can be interpreted as strong instruments in the NPIV model. Contrarily, instruments are then weak if $\text{diam}(I_0(P)) \geq \varepsilon$. Such a definition of weak instruments is related to the definition of Stock and Yogo (2005) in the linear model who also think of weak instruments in terms of properties of estimators.

In this paper I show that, under certain assumptions, the null hypothesis

$$H_0 : \text{diam}(I_0(P)) \geq \varepsilon$$

is testable. By the previous arguments rejecting H_0 provides evidence for both a small identified set and a small asymptotic bias of estimators. Notice that either H_0 is true or $\|\tilde{g} - g_0\|_c \leq \varepsilon + o_p(1)$, which allows me to link the test outcome to properties of an estimator. Specifically, I provide a test statistic, a critical value, and an estimator $\hat{g}$ such that uniformly over all distributions satisfying Assumption 1 below

$$P\left(\|\hat{g} - g_0\|_c \geq \varepsilon, \text{ reject } H_0\right) \to 0,$$

even as $\varepsilon \to 0$. I also show that the test rejects with probability approaching 1 for a large class of complete distributions and certain sequences of incomplete distributions.

To construct a test statistic, notice that if $\text{diam}(I_0(P)) \geq \varepsilon$, then there exist $g_1 \in I_0(P)$ and $g_2 \in I_0(P)$ such that $\|g_1 - g_2\|_c \geq \varepsilon$. Let $g = g_1 - g_2$. Then $E(g(X) \mid Z = z) = 0$ almost surely, $\|g\|_s \leq 2C$, and $\|g\|_c \geq \varepsilon$. Next rewrite

$$\begin{aligned} E(g(X) \mid Z = z) = 0 \text{ a.s.} \quad &\Leftrightarrow \quad E(g(X) \mid Z = z) f_Z(z) = 0 \text{ a.s.} \\ &\Leftrightarrow \quad \int \left(E(g(X) \mid Z = z) f_Z(z)\right)^2 dz = 0 \\ &\Leftrightarrow \quad \int \left(\int g(x) f_{XZ}(x,z) dx\right)^2 dz = 0 \end{aligned}$$

and define

$$S_0(g) \equiv \int \left(\int g(x) f_{XZ}(x,z) dx\right)^2 dz$$

and

$$T \equiv \inf_{g : \|g\|_s \leq 2C, \|g\|_c \geq \varepsilon} S_0(g).$$

If H_0 is true, then $T = 0$. Moreover, under the assumptions below, $T > 0$ for certain alternatives, among others all complete distributions (see Theorem 1 for details). Also notice that with $C = \infty$, T would be equal to 0 for both complete and incomplete distributions and thus, imposing smoothness restrictions on g_0 is critical.

Finally notice that the infimum will be attained at a function where $\|g\|_c = \varepsilon$, because otherwise we could simply scale down g. Moreover,

$$\inf_{g : \|g\|_s \leq 2C, \|g\|_c = \varepsilon} S_0(g) = \inf_{g : \|g/\varepsilon\|_s \leq 2C/\varepsilon, \|g/\varepsilon\|_c = 1} \varepsilon^2 S_0(g/\varepsilon) = \inf_{g \in \bar{\mathcal{G}}(\varepsilon) : \|g\|_c = 1} \varepsilon^2 S_0(g),$$

If ε changes with the sample size, then the function space changes with the sample size as well. Neglecting ε^2 in front of the objective does not change the minimizer, so I will consider a test statistic based on a scaled sample analog of $\inf_{g\in\bar{\mathcal{G}}(\varepsilon):\|g\|_c=1} S_0(g)$.

3 Estimation and testing

I now present the sample analog of T, the estimator of g_0, and the main results which, among others, link the outcome of the test to properties of the estimator. Throughout the paper I assume that the data is a random sample of (Y, X, Z), where X and Z are continuously distributed scalar random variables with compact support and joint density f_{XZ}. We can then assume without loss of generality that $X, Z \in [0, 1]$.[5]

3.1 Sample analog of test statistic

Let ϕ_j be an orthonormal basis for functions in $L^2[0, 1]$. Denote the series approximation of f_{XZ} by

$$f_J(x, z) = \sum_{j=1}^{J}\sum_{k=1}^{J} a_{jk}\phi_j(z)\phi_k(x),$$

where $a_{jk} = \int\int \phi_k(x)\phi_j(z)f_{XZ}(x, z)dxdz$. Hence, f_J is the L^2 projection onto the space spanned by the basis functions. We can then estimate f_{XZ} by

$$\hat{f}_{XZ}(x, z) = \sum_{j=1}^{J}\sum_{k=1}^{J} \hat{a}_{jk}\phi_j(z)\phi_k(x),$$

where $J \to \infty$ as $n \to \infty$ and

$$\hat{a}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \phi_j(Z_i)\phi_k(X_i).$$

Denote the series approximation of a function g by

$$g_J(x) = \sum_{j=1}^{J} h_j\phi_j(x),$$

[5]Section S.4 in the supplement outlines extensions to vectors X and Z and functions on $\mathbb{R}$.

where $h_j = \int g(x)\phi_j(x)dx \in \mathbb{R}$ for all $j = 1, \ldots, J$. Define the sieve space

$$\bar{\mathcal{G}}_J(\varepsilon) = \left\{ g \in \bar{\mathcal{G}}(\varepsilon) : g(x) = \sum_{j=1}^{J} h_j \phi_j(x) \text{ for some } h \in \mathbb{R}^J \right\}.$$

We can now define the test statistic which is

$$\hat{T} = \inf_{g \in \bar{\mathcal{G}}_J(\varepsilon) : \|g\|_c = 1} \int \left(\int g(x) \hat{f}_{XZ}(x,z) dx \right)^2 dz.$$

To obtain a simpler representation of the test statistic, let $\hat{A}$ be the $J \times J$ matrix with elements $\hat{a}_{jk}$ and let A be the population analog. Let h be the $J \times 1$ vector containing the coefficients h_j of $g \in \bar{\mathcal{G}}_J(\varepsilon)$. It is easy to show that

$$\int \left(\int g(x) \hat{f}_{XZ}(x,z) dx \right)^2 dz = \sum_{j=1}^{J} \left(\sum_{k=1}^{J} \hat{a}_{jk} h_k \right)^2 = \left\| \hat{A} h \right\|^2 = h'(\hat{A}'\hat{A})h.$$

Hence

$$\hat{T} = \inf_{g \in \bar{\mathcal{G}}_J(\varepsilon) : \|g\|_c = 1} h'(\hat{A}'\hat{A})h.$$

$\hat{T}$ depends on $\|\cdot\|_c$ and $\|\cdot\|_s$, but as shown in the next section using specific norms, it has an intuitive interpretation as a constrained version of a rank test of $A'A$.

3.2 Interpretation of test statistic with Sobolev spaces

As a particular example let

$$\|g\|_c^2 = \int_0^1 g(x)^2 dx \quad \text{and} \quad \|g\|_s^2 = \int_0^1 \left(g(x)^2 + g'(x)^2 \right) dx.$$

Furthermore, define $b_{jk} = \int \phi_j'(x)\phi_k'(x)dx$ and B as the $J \times J$ matrix with element (j,k) equal to b_{jk}. It is then easy to show that

$$\left\{ g \in \bar{\mathcal{G}}_J(\varepsilon) : \|g\|_c = 1 \right\} = \left\{ g_J : h'Bh \leq (2C/\varepsilon)^2 - 1,\ h'h = 1 \right\}.$$

It follows that the test statistic is the solution to

$$\begin{aligned} \min_{h \in \mathbb{R}^J} \quad & h'(\hat{A}'\hat{A})h \\ \text{subject to} \quad & h'Bh \leq (2C/\varepsilon)^2 - 1 \quad \text{and} \quad h'h = 1. \end{aligned}$$

Without the first constraint, the solution is the smallest eigenvalue of $\hat{A}'\hat{A}$, which could be used to test the rank of $A'A$ if J was fixed (see for example Robin and Smith, 2000). Thus, the test in this paper can be interpreted as a constrained version of a rank test, where the dimension of the matrix increases with the sample size.

3.3 Estimator

The estimator I use is a series estimator from Horowitz (2012). To describe the estimator, let $\hat{m}$ be a $J \times 1$ vector with $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_k(Z_i)$. Let

$$\hat{h} = \underset{h \in \mathbb{R}^J : \|g_J\|_s \leq C}{\arg\min} \left\| \hat{A}h - \hat{m} \right\|^2 \quad \text{and} \quad \hat{g}(x) = \sum_{j=1}^{J} \hat{h}_j \phi_j(x).$$

Without the constraint $\|g_J\|_s \leq C$ and if $\hat{A}$ is invertible, it can be shown that

$$\hat{h} = \hat{A}^{-1}\hat{m} = \left(\Phi(Z)'\Phi(X)\right)^{-1} \Phi(Z)'Y,$$

where $\Phi(W)$ is a $n \times J$ matrix containing $\phi_j(W_i)$ and Y is the $n \times 1$ vector containing Y_i. Hence, the estimator is a constrained version of the "just identified" two stage least squares estimator. In Section S.4.1 of the supplement I show that the results can easily be extended to an "over-identified" setting and non-scalar random variables.

3.4 Assumptions and main results

I will next state and discuss the assumptions and the main results.

Assumption 1. The data $\{Y_i, X_i, Z_i\}_{i=1}^{n}$ is an iid sample from the distribution of (Y, X, Z), where (Y, X, Z) are continuously distributed, $(X, Z) \in [0,1]^2$, $0 < f_{XZ}(x,z) \leq C_d < \infty$ almost everywhere, and $E(Y^2 \mid Z) \leq \sigma_Y^2$ for some $\sigma_Y > 0$. For some $r > 0$, $\|f_{XZ} - f_J\|_2 \leq C_f J^{-r}$. The data is generated by model (1) and $\|g_0\|_s \leq C$ for some constant $C > 0$.

Let $\mathcal{P}$ be the class of distributions P satisfying Assumption 1. For a fixed $\varepsilon > 0$, define $\mathcal{P}_0$ and $\mathcal{P}_1$ as the distributions in $\mathcal{P}$ satisfying H_0 and $H_1 : \text{diam}(I_0(P)) < \varepsilon$, respectively. The remaining assumptions are as follows.

Assumption 2. $\mathcal{G}$ is compact under $\|\cdot\|_c$ and $C_o\|g\|_c^2 \geq \|g\|_2^2$ for some $C_o > 0$.

Assumption 3. The basis functions form an orthornomal basis of $L^2[0,1]$.

Assumption 4. For all $g \in \mathcal{G}$ and some $C_b > 0$, $\|g - g_J\|_c \leq C_b J^{-\bar{s}}$ with $\bar{s} \geq 2$.

Assumption 5. For all $g \in \mathcal{G}$ and for J large enough, $g_J \in \mathcal{G}$ and $\frac{\|g\|_c}{\|g_J\|_c} g_J \in \mathcal{G}$.

The first assumption restricts the class of distributions. Compactness in Assumption 2 is implied by many standard choices of norms, among other for the norms used in Section 3.2 (see Appendix S.3.1 for more details). The second part of Assumption 2 implies that $S_0(g)$ is continuous in g under $\|\cdot\|_c$. It allows $\|\cdot\|_c$ to be the L^2-norm, the sup-norm, and many other norms. Assumptions 3 and 4 are standard in the literature. Assumption 5 implies that the series approximations of functions in $\bar{\mathcal{G}}(\varepsilon)$ are in $\bar{\mathcal{G}}_J(\varepsilon)$ and are therefore contained in the set that is minimized over in the definition of the test statistic. It is stronger than necessary and it can be relaxed at the expense of additional notation, but it appears to be reasonable as argued in Appendix S.3. Finally, to state the main result let

$$\kappa(P,\varepsilon) = \inf_{g \in \bar{\mathcal{G}}(\varepsilon): \|g\|_c = 1} \int \left(\int g(x) f_{XZ}(x,z) dx \right)^2 dz.$$

Notice that for any fixed distribution for which there is no $g \in \bar{\mathcal{G}}(\varepsilon)$ with $\|g\|_c = 1$ and $S_0(g) = 0$ (e.g. any complete distribution), it holds by Assumption 2 that $\kappa(P,\varepsilon) > 0$. We now get the following result. All proofs are in Appendix A.

Theorem 1. *Suppose Assumptions 1 – 5 hold.*

1. *If* $\frac{J}{c_n} \to 0$ *and* $\frac{nJ^{-2r}}{c_n} \to 0$, *then*

$$\sup_{P \in \mathcal{P}_0} P\big(n\hat{T} \geq c_n\big) \to 0.$$

2. *If* $\frac{n}{c_n} \to \infty$ *and* $\frac{J^2}{n} \to 0$, *then for all* $P \in \mathcal{P}_1$ *with* $\kappa(P,\varepsilon) > 0$,

$$P\big(n\hat{T} \geq c_n\big) \to 1.$$

3. *If* $\frac{J}{c_n} \to 0$ *and* $\frac{nJ^{-2\bar{s}}}{c_n} \to 0$, *then*

$$\sup_{P \in \mathcal{P}} P\big(\|\hat{g} - g_0\|_c \geq \varepsilon,\, n\hat{T} \geq c_n\big) \to 0.$$

The first and the second part of Theorem 1 show that $H_0 : \text{diam}(I_0(P)) \geq \varepsilon$ is a testable hypothesis. Here I focus on a diverging critical value c_n under simple assumptions in order to obtain the main result that links the test outcome to properties of $\hat{g}$. The third part implies that for any sequence of distributions $P_n \in \mathcal{P}$ for which $P_n(n\hat{T} \geq c_n) \geq \delta > 0$, it holds that

$$P_n\left(\|\hat{g} - g_0\|_c \geq \varepsilon \mid n\hat{T} \geq c_n\right) \to 0.$$

Hence, rejecting H_0 provides evidence for a small asymptotic bias and the second part of Theorem 1 shows that the rejection probability converges to 1 for a certain class of distributions, which includes all fixed complete distributions. While the results rely on a particular estimator and test statistic, the main idea extends to alternative estimator and test pairs.

The rate conditions of parts 1 - 3 of Theorem 1 are satisfied if

$$\left(\frac{J}{c_n}, \frac{c_n}{n}\right) \to 0 \qquad \text{and} \qquad \left(\frac{n}{J^{2r+1}}, \frac{n}{J^{2\bar{s}+1}}, \frac{J^2}{n}\right) \to 0.$$

Hence, both c_n and J have to go to ∞ but cannot diverge too fast relative to n. If $r \geq 2$, feasible choices are $c_n = J\ln(n)$ and $J = n^a$, where $a \in (1/5, 1/2)$.

Remark 1. Theorem A1 in Appendix S.1 demonstrates that the results in Theorem 1 also hold when $\varepsilon \to 0$ as $n \to \infty$. Therefore, the data can provide empirical evidence for an arbitrary small identified set and an arbitrary small asymptotic bias of the estimator. Canay et al. (2013) prove that the data cannot provide conclusive evidence in favor of completeness because for any complete distribution, there is a sequence of incomplete distributions, which converges to it in the total variation distance. Hence, any test that controls size over every sequence of incomplete distributions has power no larger than size against any alternative. For the test discussed here $P_n(n\hat{T} \geq c_n) \to 1$ for fixed complete distributions even as $\varepsilon \to 0$ and therefore, it cannot control size uniformly over all incomplete distributions. However, the results in Theorem A1 show that for certain sequences of incomplete distributions, where size is not controlled, $\|\hat{g} - g_0\|_c \xrightarrow{p} 0$. Thus, for n large enough either the test does not reject or $\hat{g}$ is arbitrarily close to g_0 with probability arbitrarily close to 1.

3.5 Estimating the diameter of the identified set

Let c_n be a sequence of critical values that does not depend on ε and let

$$\hat{T}(\varepsilon) = \inf_{g \in \bar{\mathcal{G}}_J(\varepsilon): \|g\|_c = 1} \int \left(\int g(x)\hat{f}_{XZ}(x,z)dx\right)^2 dz$$

for all $\varepsilon > 0$ and $\hat{T}(0) = 0$. Notice that $\hat{T}(\varepsilon)$ is increasing in ε. Define

$$\hat{\varepsilon} = \sup\left\{\varepsilon \in [0, \tilde{C}] : n\hat{T}(\varepsilon) \leq c_n\right\},$$

where $\tilde{C}$ is the largest ε such that $\{g : \|g\|_s \leq 2C, \|g\|_c \geq \varepsilon\} \neq \emptyset$. Let $\varepsilon_0 = \text{diam}(I_0(P))$ and let $\tilde{\varepsilon}_0 = \sup\{\varepsilon \in [0, \tilde{C}] : \inf_{g:\|g\|_s \leq 2C, \|g\|_c = \varepsilon} S_0(g) = 0\}$. By construction, $\varepsilon_0 \leq \tilde{\varepsilon}_0$ and for any fixed complete distribution $\varepsilon_0 = \tilde{\varepsilon}_0 = 0$. We now get the following result.[6]

Theorem 2. *Suppose Assumptions 1 - 5 hold.*

1. *If* $\frac{J}{c_n} \to 0$ *and* $\frac{nJ^{-2r}}{c_n} \to 0$*, then* $\sup_{P \in \mathcal{P}} |P(\hat{\varepsilon} \geq \varepsilon_0 - \gamma) - 1| \to 0$ *for any* $\gamma > 0$.
2. *If* $\frac{n}{c_n} \to \infty$ *and* $\frac{J^2}{n} \to 0$*, then* $P(\hat{\varepsilon} \leq \tilde{\varepsilon}_0 + \gamma) \to 1$ *for any* $\gamma > 0$.

Hence, $\hat{\varepsilon}$ is a uniformly consistent estimated upper bound of ε_0. The second result does not hold uniformly over all distributions in $\mathcal{P}$. The reason is that there are sequences of complete distributions, with $\tilde{\varepsilon}_0 = \varepsilon_0 = 0$, which are arbitrarily close to incomplete distributions with $\tilde{\varepsilon}_0 > \varepsilon_0 > 0$. For such sequences, the data cannot distinguish between the complete and the incomplete distributions and thus $\hat{\varepsilon}$ can be large even if $\tilde{\varepsilon}_0 = \varepsilon_0 = 0$. For such sequences also $\hat{g}$ would not be a consistent estimator of g_0 and therefore, the test described above would not reject $H_0 : \text{diam}(I_0(P)) \geq \varepsilon$ for small $\varepsilon > 0$. However, a consequence of the theorem is that for any fixed complete distribution it holds that $\hat{\varepsilon} \xrightarrow{p} 0$.

For incomplete distributions, $\hat{\varepsilon}$ generally does not converge in probability to ε_0, but is contained in $[\varepsilon_0 - \gamma, \tilde{\varepsilon}_0 + \gamma]$ for any $\gamma > 0$ with probability approaching 1. The discrepancy between the lower bound and the upper bound comes from the fact that the test is based on a certain implication of H_0, as explained in Section 2.2.

Remark 2. Suppose $\|\cdot\|_c$ is the sup norm. Let $\tilde{g}$ be any estimator of g_0 such that

$$\inf_{g \in I_0(P)} \|\tilde{g} - g\|_c = o_p(1).$$

and let $\tilde{g}_l(x) = \tilde{g}(x) - \hat{\varepsilon}$ and $\tilde{g}_u(x) = \tilde{g}(x) + \hat{\varepsilon}$. It is easy show that under the assumptions of Theorem 2 for any $\gamma > 0$

$$P(\tilde{g}_l(x) - \gamma \leq g(x) \leq \tilde{g}_u(x) + \gamma \text{ for all } x \in [0,1] \text{ and all } g \in I_0(P)) \;\; \to \;\; 1.$$

[6] I thank an anonymous referee for a suggestion that led to this result.

Thus, $[\tilde{g}_l(x), \tilde{g}_u(x)]$ is a set estimator of the identified set, which is easy to report and arbitrarily close to g_0 for complete distributions if n is large enough. For incomplete distributions it is generally a conservative estimator of the identified set. Section S.2 in the supplement briefly describes how a diverging c_n can be chosen using an increasing quantile of a bootstrap distribution. Hence, just like other estimators of identified sets (e.g. Chernozhukov et al. (2007) and Santos (2012)), it requires a tuning parameter, and the one presented here has a clear interpretation for a given choice.

4 Discussion and conclusion

This paper shows that even though the data cannot provide evidence in favor of completeness, it can provide evidence in favor of both an arbitrarily small identified set and an arbitrarily small asymptotic bias of an estimator in the NPIV model. The results can be used to estimate an upper bound of the diameter of the identified set and to obtain an easy to report estimator of the identified set itself.

Santos (2012) provides pointwise valid confidence sets for functionals of g_0, which are robust to partial identification. These sets are obtained by test inversion, which means that a hypothesis test has to be performed for each possible value of the functional and the bootstrap critical value depends on the hypothesized value of the functional. Hence, such a confidence set is computationally expensive to obtain. The results in this paper could potentially also be used to obtain a simple confidence set for the entire function g_0, rather than a functional, which is robust to partial identification and easy to report. In particular, the estimator of Chen and Pouzo (2012) is consistent for the function in the identified set with the minimal norm (which is identified), denoted by g_m. Now suppose there exists an estimated upper bound function $\tilde{g}_u$ and a lower bound function $\tilde{g}_l$ such that

$$P\left(\tilde{g}_l(x) \leq g_m(x) \leq \tilde{g}_u(x) \text{ for all } x \in [0,1]\right) \to 1-\alpha.$$

Similar as in Remark 2, it is then easy to show that under certain assumptions

$$\liminf_{n\to\infty} P\left(\tilde{g}_l(x) - \hat{\varepsilon} \leq g(x) \leq \tilde{g}_u(x) + \hat{\varepsilon} \text{ for all } x \in [0,1] \text{ and all } g \in I_0(P)\right) \geq 1-\alpha.$$

That is, a uniform confidence band for the identified function g_m could easily be transformed to a confidence set for the identified set using the results in this paper.

A Proofs of Theorems 1 and 2

A.1 Proof of Theorem 1

For all $P \in \mathcal{P}_0$, there exists a function g with $S_0(g) = 0$, $\|g\|_c = 1$ and $\|g\|_s \leq (2C/\varepsilon)$. Let g_J be the series approximation of such a function. Assumption 5 implies that $g_J/\|g_J\|_c \in \bar{\mathcal{G}}_J(\varepsilon)$. Let $h \in \mathbb{R}^J$ be the vector containing the coefficients of this normalized series approximation. Then

$$n\|\hat{A}h\|^2 = n\|(\hat{A} - A)h + Ah\|^2 \leq 2\|\sqrt{n}(\hat{A} - A)h\|^2 + 2n\|Ah\|^2.$$

Notice that

$$\begin{aligned}
\|Ah\|^2 &= \frac{1}{\|g_J\|_c^2} \int \left(\int f_J(x,z) g_J(x) dx \right)^2 dz \\
&= \frac{1}{\|g_J\|_c^2} \int \left(\int f_J(x,z) g(x) dx \right)^2 dz \\
&= \frac{1}{\|g_J\|_c^2} \int \left(\int (f_J(x,z) - f_{XZ}(x,z)) g(x) dx \right)^2 dz \\
&\leq \frac{1}{\|g_J\|_c^2} \left(\int \int (f_J(x,z) - f_{XZ}(x,z))^2 dx dz \right) \left(\int g(x)^2 dx \right) \\
&\leq \frac{1}{\|g_J\|_c^2} C_f^2 C_o J^{-2r}.
\end{aligned}$$

Also notice that by Assumption 4, $|\|g_J\|_c - 1| \leq 2C_b J^{-\bar{s}}/\varepsilon$. Next write

$$\|\sqrt{n}(\hat{A} - A)h\|^2 = \sum_{j=1}^J \left(\sum_{k=1}^J \sqrt{n}(\hat{a}_{jk} - a_{jk}) h_k \right)^2$$

and notice that

$$\sum_{k=1}^J \sqrt{n}(\hat{a}_{jk} - a_{jk}) h_k = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^J (\phi_k(X_i)\phi_j(Z_i) - E(\phi_k(X_i)\phi_j(Z_i))) h_k$$

and

$$Var\left(\sum_{k=1}^J (\phi_k(X_i)\phi_j(Z_i) - E(\phi_k(X_i)\phi_j(Z_i))) h_k \right) \leq \max_{k=1,\ldots,J} E(\phi_k(X_i)^2 \phi_j(Z_i)^2) \left(\sum_{k=1}^J |h_k| \right)^2.$$

Assumptions 2 and 4 imply that for some constant $\bar{C}$, $|h_k| \leq \frac{\bar{C}}{k^2\varepsilon}$ (see Appendix S.3 for a derivation). Moreover, by Assumptions 1 and 3, $E(\phi_k(X_i)^2\phi_j(Z_i)^2) \leq C_d$. It follows that

$$Var\left(\sum_{k=1}^{J}(\phi_k(X_i)\phi_j(Z_i) - E(\phi_k(X_i)\phi_j(Z_i)))h_k\right) \leq \sigma^2/\varepsilon^2,$$

where $\sigma^2 = C_d\bar{C}^2\left(\sum_{k=1}^{\infty}k^{-2}\right)^2 < \infty$. By Markov's inequality and for all n large enough

$$\begin{aligned}
\sup_{P\in\mathcal{P}_0} P\left(n\hat{T} \geq c_n\right) &\leq \sup_{P\in\mathcal{P}_0} P\left(2\|\sqrt{n}(\hat{A}-A)h\|^2 + 4C_f^2C_o nJ^{-2r} \geq c_n\right) \\
&\leq \sup_{P\in\mathcal{P}_0} P\left(\|\sqrt{n}(\hat{A}-A)h\|^2 \geq \frac{1}{4}c_n\right) \\
&\leq \frac{4J\sigma^2}{\varepsilon^2 c_n} \\
&\to 0.
\end{aligned}$$

For the second part, for any $g \in \bar{\mathcal{G}}_J(\varepsilon)$ with $\|g\|_c = 1$, let $h \in \mathbb{R}^J$ be the coefficients of the series expansion and notice that $\|\hat{A}h\|^2 \geq \frac{3}{4}\|Ah\|^2 - 3\|(\hat{A}-A)h\|^2$.[7] Moreover

$$\begin{aligned}
\|Ah\|^2 &= \int\left(\int f_J(x,z)g(x)dx\right)^2 dz \\
&= \int\left(\int f_{XZ}(x,z)g(x)dx + \int(f_J(x,z) - f_{XZ}(x,z))g(x)dx\right)^2 dz \\
&= \int\left(\int f_{XZ}(x,z)g(x)dx\right)^2 dz + \int\left(\int(f_J(x,z) - f_{XZ}(x,z))g(x)dx\right)^2 dz \\
&\quad +2\int\left(\int(f_J(x,z) - f_{XZ}(x,z))g(x)dx\right)\left(\int f_{XZ}(x,z)g(x)dx\right)dz
\end{aligned}$$

and therefore, by the Cauchy-Schwarz inequality, for all $g \in \bar{\mathcal{G}}_J(\varepsilon)$ with $\|g\|_c = 1$

$$\left|\|Ah\|^2 - \int\left(\int f_{XZ}(x,z)g(x)dx\right)^2 dz\right| \leq C_f^2C_oJ^{-2r} + 2C_dC_fC_oJ^{-r}.$$

It follow that

$$\inf_{g\in\bar{\mathcal{G}}_J(\varepsilon):\|g\|_c=1}\|Ah\|^2 = \inf_{g\in\bar{\mathcal{G}}_J(\varepsilon):\|g\|_c=1} S_0(g) + o(1) \geq \kappa(P,\varepsilon) + o(1).$$

[7] For $a = \hat{A}h$ and $b = Ah$ it holds that $0 \leq \|2a - \frac{3}{2}b\|^2 = 3\|a-b\|^2 - \frac{3}{4}\|b\|^2 + \|a\|^2$.

For $\kappa(P,\varepsilon) > 0$ and n large enough $n\|Ah\|^2 \geq \frac{1}{3}n\kappa(P,\varepsilon)$ and $c_n \leq \frac{1}{8}n\kappa(P,\varepsilon)$. Also

$$\|(\hat{A}-A)h\|^2 = \sum_{j=1}^{J}\left(\sum_{k=1}^{J}(\hat{a}_{jk}-a_{jk})h_k\right)^2 \leq C_o \sum_{j=1}^{J}\sum_{k=1}^{J}(\hat{a}_{jk}-a_{jk})^2$$

for all h with $\|h\|^2 \leq C_o$. Finally, notice that

$$\sqrt{n}(\hat{a}_{jk}-a_{jk}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\phi_k(X_i)\phi_j(Z_i) - E(\phi_k(X_i)\phi_j(Z_i)))$$

and $Var\left(\phi_k(X_i)\phi_j(Z_i) - E(\phi_k(X_i)\phi_j(Z_i))\right) \leq C_d$. Now for all n large enough

$$\begin{aligned} P\left(n\hat{T} \geq c_n\right) &\geq P\left(\frac{1}{4}n\kappa(P,\varepsilon) - 3C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk}-a_{jk}))^2 \geq c_n\right) \\ &\geq P\left(\frac{1}{8}n\kappa(P,\varepsilon) \geq 3C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk}-a_{jk}))^2\right) \\ &= 1 - P\left(24C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk}-a_{jk}))^2 \leq n\kappa(P,\varepsilon)\right) \\ &\geq 1 - \frac{24J^2C_oC_d}{n\kappa(P,\varepsilon)}. \end{aligned}$$

Hence $P(n\hat{T} \geq c_n) \to 1$.

For the last claim let $h_0 \in \mathbb{R}^J$ contain the first J coefficients of the series expansion of g_0. By Assumption 5, the definition of $\hat{h}$ and the triangle inequality

$$\left\|\hat{A}h_0 - \hat{m}\right\| \geq \left\|\hat{A}\hat{h} - \hat{m}\right\| = \left\|\hat{A}h_0 - \hat{m} + \hat{A}(\hat{h}-h_0)\right\| \geq \left\|\hat{A}(\hat{h}-h_0)\right\| - \left\|\hat{A}h_0 - \hat{m}\right\|$$

and thus

$$4\left\|\hat{A}h_0 - \hat{m}\right\|^2 \geq \left\|\hat{A}(\hat{h}-h_0)\right\|^2.$$

Now suppose that $n\hat{T} \geq c_n$ and $\|\hat{g}-g_0\|_c \geq \varepsilon$ and notice that $\|\hat{g}-g_0\|_s \leq 2C$. Let $\tilde{g} = \hat{g} - g_0$. From Assumption 5 it follows that $\tilde{g}_J\frac{\|\tilde{g}\|_c}{\varepsilon\|\tilde{g}_J\|_c} \in \bar{\mathcal{G}}_J(\varepsilon)$. Therefore,

$$\frac{n}{\varepsilon^2}\frac{\|\tilde{g}\|_c^2}{\|\tilde{g}_J\|_c^2}\left\|\hat{A}(\hat{h}-h_0)\right\|^2 \geq n\hat{T} \geq c_n.$$

By Assumption 4 and $\|\tilde{g}\|_c \geq \varepsilon$, we have $\left|\frac{\|\tilde{g}_J\|_c}{\|\tilde{g}\|_c} - 1\right| \leq 2C_bJ^{-\bar{s}}/\varepsilon$ and thus

$$8\frac{n}{\varepsilon^2}\left\|\hat{A}h_0 - \hat{m}\right\|^2 \geq c_n$$

for n large enough. In other words, for n large enough

$$\sup_{P\in\mathcal{P}} P\left(\|\hat{g}-g_0\|_c \geq \varepsilon, n\hat{T} > c_n\right) \;\leq\; \sup_{P\in\mathcal{P}} P\left(8n\left\|\hat{A}h_0-\hat{m}\right\|^2 \geq c_n\varepsilon^2\right).$$

Next let m be a $J\times 1$ vector with $m_k = E(\hat{m}_k)$ and notice that since

$$\left\|\hat{A}h_0-\hat{m}\right\| \leq \|(\hat{A}-A)h_0\| + \|Ah_0-m\| + \|m-\hat{m}\|$$

we have

$$\left\|\hat{A}h_0-\hat{m}\right\|^2 \leq 4\|(\hat{A}-A)h_0\|^2 + 4\left\|Ah_0-m\right\|^2 + 4\left\|m-\hat{m}\right\|^2$$

and thus for n large enough

$$\begin{aligned} &P\left(8n\left\|\hat{A}h_0-\hat{m}\right\|^2 \geq c_n\varepsilon^2\right) \\ &\qquad \leq P\left(32n\|(\hat{A}-A)h_0\|^2 + 32n\left\|Ah_0-m\right\|^2 + 32n\left\|m-\hat{m}\right\|^2 \geq c_n\varepsilon^2\right) \\ &\qquad \leq P\left(96n\|(\hat{A}-A)h_0\|^2 \geq c_n\varepsilon^2\right) + P\left(96n\left\|Ah_0-m\right\|^2 \geq c_n\varepsilon^2\right) \\ &\qquad\quad +P\left(96n\left\|m-\hat{m}\right\|^2 \geq c_n\varepsilon^2\right). \end{aligned}$$

It now suffices to prove that all three terms on the right hand side converge to 0 uniformly over $P\in\mathcal{P}$.

The first term converges to 0 using arguments identical to the ones in the proof of the first part. Similarly, for the third term

$$\|\sqrt{n}(m-\hat{m})\|^2 = \sum_{k=1}^{J}\left(\sqrt{n}(\hat{m}_k-m_k)\right)^2 = \sum_{k=1}^{J}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i\phi_k(Z_i)-E(Y_i\phi_k(Z_i))\right)\right)^2$$

and by Assumptions 1 and 3

$$E\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i\phi_k(Z_i)-E(Y_i\phi_k(Z_i))\right)\right)^2\right) \leq \frac{1}{n}\sum_{i=1}^{n}E\left(Y_i^2\phi_k(Z_i)^2\right) \leq \sigma_Y^2 C_d.$$

It follows from Markov's inequality that

$$\sup_{P\in\mathcal{P}} P\left(96n\|m-\hat{m}\|^2 \geq c_n\varepsilon^2\right) \leq \frac{96J\sigma_Y^2C_d}{c_n\varepsilon^2} \to 0.$$

Finally write

$$\|Ah_0-m\|^2 = \sum_{j=1}^{J}\left(\sum_{k=1}^{J}a_{jk}h_{0,k}-m_j\right)^2.$$

Since $\sum_{j=1}^{\infty}\left(\sum_{k=1}^{\infty} a_{jk}h_{0,k} - m_j\right)^2 = 0$ it holds that $m_j = \sum_{k=1}^{\infty} a_{jk}h_{0,k}$. Therefore

$$\begin{aligned}
\|Ah_0 - m\|^2 &= \sum_{j=1}^{J}\left(\sum_{k=1}^{J} a_{jk}h_{0,k} - m_j\right)^2 \\
&\leq \sum_{j=1}^{\infty}\left(\sum_{k=J+1}^{\infty} a_{jk}h_{0,k}\right)^2 \\
&= \int\left(\int f_{XZ}(x,z)\left(g_0(x) - g_{0,J}(x)\right)dx\right)^2 dz \\
&\leq \int\int f_{XZ}(x,z)^2 dx dz \int \left(g_0(x) - g_{0,J}(x)\right)^2 dx \\
&\leq C_d^2 C_b^2 C_o J^{-2\bar{s}},
\end{aligned}$$

where the last inequality follows from Assumptions 2 and 4. Thus,

$$\sup_{P\in\mathcal{P}} P\left(96n\,\|Ah_0 - m\|^2 \geq c_n\varepsilon^2\right) \leq \sup_{P\in\mathcal{P}} P\left(96C_d^2C_b^2C_oJ^{-2\bar{s}}n \geq c_n\varepsilon^2\right) \to 0.$$

We can conclude that $\sup_{P\in\mathcal{P}} P(\|\hat{g} - g_0\|_c \geq \varepsilon, n\hat{T} > c_n) \to 0$.

A.2 Proof of Theorem 2

I show that $\sup_{P\in\mathcal{P}} P(\hat{\varepsilon} < \varepsilon_0 - \gamma) \to 0$ and $P(\tilde{\varepsilon}_0 + \gamma < \hat{\varepsilon}) \to 0$ for any $\gamma > 0$.

If $\hat{\varepsilon} < \varepsilon_0 - \gamma < \varepsilon_0$, by definition it holds that $n\hat{T}(\varepsilon_0) \geq c_n$ and $\varepsilon_0 \geq \gamma$. Since $\text{diam}(I_0) = \varepsilon_0$, there exists a function g with $\|g\|_s \leq 2C/\varepsilon_0$, $\|g\|_c = 1$, and $S_0(g) = 0$. Thus, the arguments of the proof of the first part of Theorem 1 imply that

$$\sup_{P\in\mathcal{P}} P(\hat{\varepsilon} < \varepsilon_0 - \gamma) \leq \sup_{P\in\mathcal{P}} P(n\hat{T}(\varepsilon_0) \geq c_n) \to 0.$$

For the second claim suppose that $\tilde{\varepsilon}_0 + \gamma < \hat{\varepsilon} \leq \tilde{C}$. Then, by definition $n\hat{T}(\tilde{\varepsilon}_0 + \gamma) \leq c_n$. Similar as in the proof of the second part of Theorem 1

$$n\hat{T}(\tilde{\varepsilon}_0 + \gamma) \;\geq\; \inf_{g\in\bar{\mathcal{G}}_J(\tilde{\varepsilon}_0+\gamma):\|g\|_c=1} \frac{3}{4}n\|Ah\|^2 - 3C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk} - a_{jk}))^2$$

and

$$\inf_{g\in\bar{\mathcal{G}}_J(\tilde{\varepsilon}_0+\gamma):\|g\|_c=1} \|Ah\|^2 \geq \inf_{g\in\bar{\mathcal{G}}(\tilde{\varepsilon}_0+\gamma):\|g\|_c=1} S_0(g) + o(1).$$

But by the definition of $\tilde{\varepsilon}_0$, it holds that $\inf_{g \in \bar{\mathcal{G}}(\tilde{\varepsilon}_0+\gamma): \|g\|_c=1} S_0(g) > 0$. Hence, using arguments analogous to those in the proof of the second part of Theorem 1, it follows that

$$P(\tilde{\varepsilon}_0 + \gamma < \hat{\varepsilon}) \leq P(n\hat{T}(\tilde{\varepsilon}_0 + \gamma) \leq c_n) \to 0.$$

References

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Andrews, D. W. K. (2011). Examples of L2-complete and boundedly-complete distributions. Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1840552.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica 75*(6), 1613–1669.

Canay, I. A., A. Santos, and A. M. Shaikh (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica 81*(6), 2535–2559.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014). Local identification of nonparametric and semiparametric models. *Econometrica 82*(2), 785–809.

Chen, X. and T. M. Christensen (2017). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. Working paper, https://arxiv.org/abs/1508.03365.

Chen, X. and D. Pouzo (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics 152*(1), 46–60.

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth moments. *Econometrica 80*(1), 277–321.

Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica 83*(3), 1013–1079.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica 79*(5), 1541–1565.

D'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory 27*(3), 460–471.

Freyberger, J. (2012). Nonparametric panel data models with interactive fixed effects. Working paper, `http://www.ssc.wisc.edu/~jfreyberger/NPPanelsFreyberger.pdf`.

Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics 33*(6), 1–27.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica 79*(2), 347–394.

Horowitz, J. L. (2012). Specification testing in nonparametric instrumental variable estimation. *Journal of Econometrics 167*(2), 383–396.

Horowitz, J. L. (2014). Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. *Journal of Econometrics 180*(2), 158–173.

Horowitz, J. L. and S. Lee (2012). Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics 168*(2), 175–188.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Hu, Y. and J.-L. Shiu (2016). Nonparametric identification using instrumental variables: sufficient conditions for completeness. Working paper, `http://www.econ2.jhu.edu/People/Hu/Identification_using_IV_2016.pdf`.

Mattner, L. (1993). Some incomplete but boundedly complete location families. *The Annals of Statistics 21*(4), 2158–2162.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Robin, J.-M. and R. J. Smith (2000). Tests of rank. *Econometric Theory 16*(2), 151–175.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica 80*(1), 213–275.

Sasaki, Y. (2015). Heterogeneity and selection in dynamic panel data. *Journal of Econometrics 188*(1), 236–249.

Stock, J. and M. Yogo (2005). Testing for weak instruments in linear IV regression. In D. Andrews and J. Stock (Eds.), *Identification and Inference for Econometric Models*. New York: Cambridge University Press.

Tao, J. (2016). On the asymptotic theory for semi-nonparametric GMM models with partial identification. Working paper, `https://docs.google.com/a/uw.edu/viewer?a=v&pid=sites&srcid=dXcuZWR1fGppbmctdGFvfGd4OjMyZThkOWJkNTNiYTljZjY`.