Supplement to "On completeness and consistency in nonparametric instrumental variable models"

Joachim Freyberger[‡]

April 14, 2017

Abstract

This supplement contains additional material to accompany the main text. First, I show that the results in Theorem 1 also hold as $\varepsilon \to 0$ and I informally outline a bootstrap procedure to select the critical value. I then provide additional explanations for the main assumptions as well as extensions of the main results.

S.1 Results when $\varepsilon \to 0$

In this section I show that the results in Theorem 1 also hold as $\varepsilon \to 0$. To do so define

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

We now obtain the following result.

Theorem A1. *Suppose Assumptions 1 – 5 hold.*

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

$$\sup_{P \in \mathcal{P}:\, \mathrm{diam}(I_0(P)) \geq \varepsilon} P\big(n\hat{T} \geq c_n\big) \to 0.$$

2. *For any* $P_n \in \mathcal{P}$ *with* $\frac{n\kappa_J(P_n,\varepsilon)}{c_n} \to \infty$ *and* $\frac{J^2}{n\kappa_J(P_n,\varepsilon)} \to 0$

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

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

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

Since $\varepsilon \to 0$, the last part now implies that for any sequence of distributions for which the test rejects with probability larger than δ, $\hat{g}$ will be consistent for g_0. To highlight that $P_n(n\hat{T} \geq c_n) \to 1$, and thus $\|\hat{g} - g_0\| \xrightarrow{p} 0$, not only for fixed complete distributions but also for certain sequences of

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

incomplete distributions, the second part is now stated in terms of sequences of distributions. A simple example of a sequence where these results hold is a sequence where the density is a series approximation f_J of a density f_{XZ} corresponding to a complete distribution. The first and third part of Theorem A1 show size control over distributions which do not yield consistency, among others all fixed distributions for which $\text{diam}(I_0(P)) > 0$.

For a fixed ε, the rate conditions of parts 1 - 3 of Theorem 1 are satisfied if

$$\text{(S1)} \qquad \left(\frac{J}{c_n}, \frac{c_n}{n}\right) \to 0 \qquad \text{and} \qquad \left(\frac{n}{J^{2r+1}}, \frac{n}{J^{2\bar{s}+1}}, \frac{J^2}{n}\right) \to 0,$$

and if $r \geq 2$ feasible choices are $c_n = J\ln(n)$ and $J = n^a$, where $a \in (1/5, 1/2)$. In this case $P(n\hat{T} \geq c_n) \to 1$ for any complete distribution. Thus, as long as ε converges to 0 slow enough, the rate conditions in (S1) together with Assumptions 1 - 5 imply that $\sup_{P\in\mathcal{P}} P(\|\hat{g} - g_0\|_c \geq \varepsilon,\, n\hat{T} \geq c_n) \to 0$ and $P(n\hat{T} \geq c_n) \to 1$ for a large class of complete distributions. The only complete distributions for which the test then does not reject with probability approaching 1 are the ones for which $a_{jk} \to 0$ as $j, k \to \infty$ extremely rapidly. For those distributions $\|\hat{g} - g_0\|_c \xrightarrow{p} 0$ very slowly and if the rate of convergence is slower than ε, the test rejects with probability approaching 0.

S.1.1 Proof of Theorem A1

For all $P \in \mathcal{P}$ with $\text{diam}(I_0(P)) \geq \varepsilon$, 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 $g_J/\|g_J\|_c \in \bar{\mathcal{G}}_J(\varepsilon)$. Now from arguments analogous to those in the proof of the first part of Theorem 1, it follows that for all n large enough

$$\begin{aligned}
\sup_{P\in\mathcal{P}_0\colon \text{diam}(I_0(P))\geq\varepsilon} P\left(n\hat{T} \geq c_n\right) &\leq \sup_{P\in\mathcal{P}_0\colon \text{diam}(I_0(P))\geq\varepsilon} P\left(2\|\sqrt{n}(\hat{A} - A)h\|^2 + 4C_f^2 C_o n J^{-2r} \geq c_n\right) \\
&\leq \sup_{P\in\mathcal{P}_0\colon \text{diam}(I_0(P))\geq\varepsilon} P\left(\|\sqrt{n}(\hat{A} - A)h\|^2 \geq \frac{1}{4}c_n\right) \\
&\leq \frac{4J\sigma^2}{\varepsilon^2 c_n}.
\end{aligned}$$

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

$$\|Ah\|^2 = \left(\int f_J(x, z) g_J(x)\right)^2 dz \geq \kappa_J(P, \varepsilon).$$

For P_n there exists n large enough such that $\frac{1}{4}n\kappa_J(P_n, \varepsilon) \geq c_n$. Thus, very similar as in the proof of Theorem 1, for all n large enough

$$\begin{aligned} P_n\left(n\hat{T} \geq c_n\right) &\geq P_n\left(\frac{3}{4}n\kappa_J(P_n,\varepsilon) - 3C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk}-a_{jk}))^2 \geq c_n\right) \\ &\geq P_n\left(\frac{1}{2}n\kappa_J(P_n,\varepsilon) \geq 3C_o\sum_{j=1}^{J}\sum_{k=1}^{J}(\sqrt{n}(\hat{a}_{jk}-a_{jk}))^2\right) \\ &\geq 1 - \frac{6J^2C_oC_d}{n\kappa_J(P_n,\varepsilon)}. \end{aligned}$$

The last part follows identical arguments as the ones in the proof of Theorem 1.

S.2 Conservative bootstrap to select c_n

In this section I informally outline a bootstrap procedure, which is conservative, but also uniformly valid, and it yields consistency against fixed alternatives under suitable assumptions. This result provides guidance about how c_n could be chosen in practice. Let $H_0 : \text{diam}(I_0(P)) \geq \varepsilon$ for a fixed $\varepsilon > 0$. 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 $\bar{h} \in \mathbb{R}^J$ be the vector containing the coefficients of this normalized series approximation and let $\hat{h} \in \mathbb{R}^J$ be the vector corresponding to the minimizer of the test statistic. Then, under assumptions similar to those in Sections 3, the arguments of the proof of Theorem 1 imply that

$$\sup_{P\in\mathcal{P}_0} P\left(n\|\hat{A}\bar{h}\|^2 \leq \bar{C}J^2\ln(\ln(n))\right) \to 1.$$

Moreover, since $n\|A\bar{h}\| \to 0$, we then get that uniformly over $P \in \mathcal{P}_0$ and for all $t > 0$

$$P\left(n\hat{T} \leq t\right) \geq P\left(n\|\hat{A}\bar{h}\|^2 \leq t\right) = P\left(n\|(\hat{A}-A)\bar{h}\|^2 \leq t\right) + o(1).$$

Let A^* be the bootstrap analog of $\hat{A}$. The proposed conservative critical value t_α is the number that satisfies

$$\min_{g_J\in\bar{\mathcal{G}}_J(\varepsilon):\|g_J\|_c=1, n\|\hat{A}h\|^2\leq n\|\hat{A}\hat{h}\|^2+\bar{C}J^2\ln(\ln(n))} P^*\left(n\|(A^*-\hat{A})h\|^2 \leq t_\alpha\right) = 1-\alpha.$$

The intuition for this procedure is that the test statistic is the minimum of some objective function over a class of functions and the population minimizer might not be unique. If we knew one of the minimizers, namely $\bar{h}$ above, then a conservative critical value could be constructed based on the $1-\alpha$ quantile of $n\|(A^*-\hat{A})\bar{h}\|^2$. While $\bar{h}$ is unknown, we know that it satisfies $n\|\hat{A}\bar{h}\|^2 \leq \bar{C}J^2\ln(\ln(n))$ with probability approaching 1. Hence, taking the largest $1-\alpha$ quantiles of $n\|(A^*-\hat{A})h\|^2$ for all suitable h such that $n\|\hat{A}h\|^2 \leq \bar{C}J^2\ln(\ln(n))$ yields a (larger) conservative critical value as well. A remaining complication is now that under a fixed alternative, there might not exist a vector h with

$n\|\hat{A}h\|^2 \leq \bar{C}J^2\ln(\ln(n))$. Contrarily, the constraint $n\|\hat{A}h\|^2 \leq n\|\hat{A}\hat{h}\|^2 + \bar{C}J^2\ln(\ln(n))$ is always satisfied for $h = \hat{h}$ and hence a feasible solution always exists. Finally notice that t_α is smaller than the $1-\alpha$ quantile of $\max_{g_J \in \bar{\mathcal{G}}_J(\varepsilon):\|g_J\|_c=1} n\|(\hat{A}-A)h\|^2$. The arguments in the proof of Theorem 1 imply that this quantile is bounded above by $t_{\alpha,u}J^2$, where $t_{\alpha,u}$ does not depend on P. These arguments also show that $n\hat{T}$ diverges at rate n when $\kappa(P,\varepsilon) > 0$. When $J^2/n \to 0$, the bootstrap critical value therefore still yields consistency against fixed alternatives with $\kappa(P,\varepsilon) > 0$.

I next outline why this critical value controls size uniformly. Since

$$\sup_{P\in\mathcal{P}_0} P\left(n\|\hat{A}\bar{h}\|^2 \leq CJ^2\ln(\ln(n))\right) \to 1,$$

it follows that

$$\sup_{P\in\mathcal{P}_0} P^*\left(n\|(A^*-\hat{A})\bar{h}\|^2 \leq t_\alpha\right) \geq 1-\alpha+o_p(1).$$

Notice that $\{M \in \mathbb{R}^{J\times J} : \|Mh\|^2 \leq t\}$ is a convex set for all $t > 0$. The results in Bentkus (2003) can therefore be used to show that under suitable assumptions

$$\sup_{P\in\mathcal{P}_0}\sup_{t\geq 0}\left|P^*\left(n\|(A^*-\hat{A})\bar{h}\|^2 \leq t\right) - P\left(n\|(\hat{A}-A)\bar{h}\|^2 \leq t\right)\right| = o_p(1).$$

It follows that uniformly over $P \in \mathcal{P}_0$

$$P\left(n\hat{T} \leq t_\alpha\right) \geq P\left(n\|(\hat{A}-A)\bar{h}\|^2 \leq t_\alpha\right) + o(1) \geq 1-\alpha+o(1)$$

and therefore

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_0} P\left(n\hat{T} \geq t_\alpha\right) \leq \alpha.$$

Finally, notice that when $\alpha \to 0$, we obtain a diverging critical value $c_n = t_\alpha$ and

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

Hence, c_n constructed in this way can be used to obtain the results in Theorem 1. While there is still ambiguity about how to choose a small α for a given sample size, the resulting critical value has an intuitive interpretation because it provides a uniformly valid upper bound for the rejection probability under the null. Moreover, the choice also automatically adapts to rescaling the data.

The critical value depends on ε, but the arguments above imply that the critical value with $\varepsilon = 0$, which can be obtained by simply dropping the constraint $g_J \in \bar{\mathcal{G}}(\varepsilon)$, is a valid and conservative critical value for all $\varepsilon > 0$. This critical value can be used for the procedure suggested in Section 3.5, which discusses the estimator $\hat{\varepsilon}$.

S.3 Additional explanations of assumptions

S.3.1 Compact function spaces

The most commonly used consistency norms are the L^2-norm, $\left(\int g(x)^2 dx\right)^{1/2}$, and the sup-norm, $\sup_x |g(x)|$. Suppose that the consistency norm is the L^2-norm. Then, as shown in Section 3.2, a convenient choice for the strong norm $\|\cdot\|_s$ is the Sobolev norm

$$\|g\|_s = \sqrt{\sum_{0\leq\lambda\leq m} \int \left(D^\lambda g(x)\right)^2 dx},$$

where $m \geq 1$ and D^λ denotes the λ weak derivative of the function $g(x)$. If instead the consistency norm is the sup-norm, one could either use the Sobolev norm above or the Hölder norm

$$\|g\|_s = \max_{0\leq|\lambda|\leq m} \sup_{x\in(0,1)} |\nabla^\lambda g(x)| + \sup_{x_1,x_2\in(0,1),x_1\neq x_2} \frac{|\nabla^m g(x_1) - \nabla^m g(x_2)|}{|x_1 - x_2|^\nu},$$

where $\nabla^\lambda g(x)$ denotes the λ derivative of the function $g(x)$, and $0 < \nu \leq 1$. In the first case $\mathcal{G}$ is a Sobolev space, while in the second case $\mathcal{G}$ is a Hölder space. Similar as the strong norm, the consistency norm could also be defined using derivatives of higher order.

In all these cases it can be shown that $\mathcal{G}$ is compact under $\|\cdot\|_c$. See Freyberger and Masten (2015) for an overview of the compactness results. Moreover, it is easy to see that with these choices $\|g\|_c^2 \geq \int g(x)^2 dx$.

S.3.2 Example of norm bounds

One assumption which is maintained throughout the paper is that $\|g_0\|_s \leq C$. As a consequence, the test involves a constraint on the strong norm and C needs be chosen by the researcher. I now briefly explain how this can be done in two popular examples, namely estimation of Engel curves and demand functions.

Let X^* be total household expenditure and let $X = \log(X^*)$. Let Y^* be the total expenditure on a certain good, such as food, and define $Y = \frac{Y^*}{X^*}$, which is the expenditure share. Let Z be the gross earnings of the head of the household. This setup is studied by Blundell et al. (2007) and Santos (2012), among others. A reasonable assumption is that if a household increases total expenditure by \$$\delta$, the total expenditure on food does not increase by more than \$$\delta$ and it does not decrease. If $X^* = \bar{X}^*$ and if we want to increase $\log(\bar{X}^*)$ to $\log(\bar{X}^*) + \delta$, then we need to increase X^* by $\bar{X}^*(\exp(\delta) - 1)$. Then the total expenditure is $\bar{X}^* \exp(\delta)$ and expenditure on food is not more than $Y^* + \bar{X}^*(\exp(\delta) - 1)$ and not less than Y^*. Therefore, it can be shown that the derivative of the Engel curve is bounded in absolute value by 1.

If $X \in [a,b]$ we can use the regressor $(X - a)/(b - a) \in [0,1]$. Then $\sup_x |g_0'(x)| \leq b - a$ and clearly $\sup_x |g_0(x)| \leq 1$. Let

$$\|g_0\|_s = \sup_{x\in(0,1)} |g_0(x)| + \sup_{x_1,x_2\in(0,1),x_1\neq x_2} \frac{|g_0(x_1) - g_0(x_2)|}{|x_1 - x_2|}.$$

Then $\|g_0\|_s \leq 1 + b - a$ and we can choose $C = 1 + b - a$. If instead

$$\|g_0\|_s = \left(\int \left(g_0(x)^2 + g_0'(x)^2\right) dx \right)^{1/2},$$

we get $\|g_0\|_s \leq \sqrt{1 + (b-a)^2}$.

If $g_0(x)$ is a demand function, then one can use bounds on price elasticities, and bounds on the support of quantity and price. In this way, one can obtain bounds on the derivatives and function values of the demand function using similar arguments as above.

S.3.3 Examples illustrating assumptions

Assumption 1 is easy to interpret, while Assumptions 2 can be verified with popular parameter spaces, including Sobolev spaces, as shown in Section S.3.1. Assumptions 3 and 4 are discussed in Chen (2007) and hold for many popular basis functions as long as sufficient smoothness is imposed. Notice that Assumptions 2 and 4 imply that for all functions $g \in \mathcal{G}$

$$\sum_{j=J+1}^{\infty} h_j^2 = \int (g(x) - g_J(x))^2 dx \leq C_o C_b^2 J^{-2\bar{s}}.$$

It follows that

$$h_{J+1}^2 \leq C_o C_b^2 J^{-2\bar{s}} \leq C_o C_b^2 2^{2\bar{s}} (J+1)^{-2\bar{s}}.$$

In other words, since the approximation error converges to 0 fast, the coefficients of the series approximation have to converge to 0 fast. Moreover, for all g with $\|g\|_s \leq (2C)/\varepsilon$, we have $\frac{\varepsilon}{2} g \in \mathcal{G}$ and thus, $\frac{\varepsilon^2}{4} h_{J+1}^2 \leq C_o C_b^2 2^{2\bar{s}} (J+1)^{-2\bar{s}}$.

Finally, to obtain more intuition for Assumption 5, suppose that $\|g_J\|_c^2 = \sum_{j=1}^J h_j^2$ and $\|g_J\|_s^2 = \sum_{j=1}^J h_j^2 (1 + b_j)$, where $b_j > 0$ is increasing and $b_j \to \infty$ as $j \to \infty$. For example, when $\|\cdot\|_s$ is the Sobolev norm, orthonormal trigonometric polynomials have this structure. Assumption 5, which could be relaxed at the expense of additional notation, says that if $\|g\|_s^2 = \sum_{j=1}^\infty h_j^2 (1 + b_j) \leq C^2$, then

$$\sum_{j=1}^{J} h_j^2 (1 + b_j) \leq C^2 \quad \text{and} \quad \frac{\sum_{j=1}^\infty h_j^2}{\sum_{j=1}^J h_j^2} \sum_{k=1}^{J} h_k^2 (1 + b_k) \leq C^2.$$

The first inequality clearly holds because $b_j > 0$. Intuitively, the series truncation leaves out the very wiggly part of g and thus, the truncation has a smaller strong norm. The second inequality says that this is true even after normalizing by the consistency norm. To see why this is true write

$$\begin{aligned}
\frac{\sum_{j=1}^{\infty} h_j^2}{\sum_{j=1}^{J} h_j^2} \sum_{j=1}^{J} h_k^2(1+b_k) &\leq \frac{\sum_{j=1}^{\infty} h_j^2 \sum_{k=1}^{J} h_k^2(1+b_k)}{\sum_{j=1}^{J} h_j^2 \sum_{k=1}^{\infty} h_k^2(1+b_k)} C^2 \\
&= \frac{\sum_{j=1}^{J}\sum_{k=1}^{J} h_j^2 h_k^2(1+b_k) + \sum_{j=J+1}^{\infty}\sum_{k=1}^{J} h_j^2 h_k^2(1+b_k)}{\sum_{j=1}^{J}\sum_{k=1}^{J} h_j^2 h_k^2(1+b_k) + \sum_{j=J+1}^{\infty}\sum_{k=1}^{J} h_j^2 h_k^2(1+b_j)} C^2 \\
&\leq C^2
\end{aligned}$$

because b_j is an increasing sequence, which implies that Assumption 5 holds.

S.4 Extensions

S.4.1 Extension to "over-identified" estimation and random vectors

As explained in Section 3.3, the estimator is a constrained version of the "just identified" two stage test squares estimator, with regressors $\phi_j(X_i)$ and instruments $\phi_j(Z_i)$ for $j = 1, \ldots, J$. Hence, the number of instruments is equal to the number of regressors. I now explain that the results can easily be extended to an "over-identified" setting. Specifically, the density could be estimated by

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

where $J \geq K$. Then $\hat{A}$ becomes a $J \times K$ matrix, $\hat{m}$ is a $J \times 1$ vector and we estimate the $K \times 1$ vector h by solving

$$\hat{h} = \underset{h \in \mathbb{R}^K : \|g_K\|_s \leq C}{\arg\min} \left\| \hat{A}h - \hat{m} \right\|_W^2,$$

where W is a positive definite weight matrix and $\|\cdot\|_W$ is the weighted Euclidean norm. The corresponding test statistic is then a scaled version of

$$\min_{h \in \mathbb{R}^K : \|g_K\|_s \leq 2C, \|g_K\|_c = \varepsilon} \|\hat{A}h\|_W^2.$$

Analogous results to the ones in this paper using identical arguments can be derived in this setting as well.

Similarly, it is easy to allow for non-scalar X and Z. Returning to the "just identified" setting with $X \in \mathbb{R}^d$ and $Z \in \mathbb{R}^d$, ϕ_j can denote basis functions for functions in $L^2[0,1]^d$. For example, when $d = 2$ and the basis functions are polynomials, $\phi_1(x)$ is the constant function, $\phi_2(x)$ a linear function of x_1, $\phi_3(x)$ a linear function of x_2, $\phi_4(x)$ a linear function of x_1x_2, etc.. Then, using the same arguments, the results in the paper still hold under the assumptions imposed.

S.4.2 Extension to functions on $\mathbb{R}$

The analysis could be extended to functions on $\mathbb{R}$ by using weighted norms. In this section, I provide the main ideas, including specific examples of norms which satisfy compactness, and the test statistic. Let $w(x) = e^{-x^2}$ and let $\phi_j(x)$ be Hermite polynomials (see for example Chen 2007) so that

$$\int \phi_j(x)^2 w(x)dx = 1$$

and for $j \neq k$

$$\int \phi_k(x)\phi_j(x)w(x)dx = 0.$$

Let the consistency norm be the weighted L^2-norm

$$\|g\|_c = \sqrt{\int g(x)^2 w(x)dx}.$$

Then for every function g for which $\int g(x)^2 w(x)dx < \infty$, we can write

$$g(x) = \sum_{j=1}^{\infty} h_j \phi_j(x),$$

where $h_j \equiv \int g(x)\phi_j(x)w(x)dx$. Moreover, if $f_{XZ}(x,z)$ is square integrable, we can write

$$f_{XZ}(x,z) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk}\phi_j(z)\phi_k(x),$$

where $a_{jk} = \int f_{XZ}(x,z)\phi_j(z)\phi_k(x)w(x)w(z)dx$. Hence, we can estimate a_{jk} by

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

and f_{XZ} by

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

Now let

$$\begin{aligned} S_0(g) &= \int \left(\int g(x) f_{XZ}(x,z)w(x)dx\right)^2 w(z)dz \\ &= \int \left(\int g(x)w(x)^{1/2} f_{XZ}(x,z)w(x)^{1/2}w(z)^{1/2}dx\right)^2 dz. \end{aligned}$$

It can be shown that $S_0(g)$ is continuous in g under $\|\cdot\|_c$ as long as $f_{XZ}(x,z)$ is bounded. With this choice of the consistency norm, we get a compact parameter space for example if

$$\|g\|_s = \sqrt{\int \sum_{0 \leq \lambda \leq m} \left(D^\lambda g(x)\right)^2 \tilde{w}(x) dx},$$

where $\tilde{w}(x) = (1 + x^2)^{-\delta}$ for any $\delta > 0$ and $D^\lambda g(x)$ denotes the λ weak derivative of g. See Freyberger and Masten (2015) for the formal compactness result, which builds on results of Gallant and Nychka (1987). With these norms we can define the parameter space just as before. Since we would assume that $\|g_0\|_s \leq C$, g_0 could be unbounded and it could have unbounded derivatives.

Due to continuity of $S_0(g)$ and compactness of the parameter space, it again holds that $\kappa(P, \varepsilon) > 0$ for every complete distribution and all $\varepsilon > 0$. This result, combined with similar assumptions as those in this paper, can be used to link the outcome of the test to properties of the estimator.

References

Bentkus, V. (2003). On the dependence of the berryesseen bound on dimension. *Journal of Statistical Planning and Inference 113*, 385–402.

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

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 76, pp. 5550–5623. Elsevier.

Freyberger, J. and M. A. Masten (2015). Compactness of infinite dimensional parameter spaces. Working paper, `www.cemmap.ac.uk/uploads/cemmap/wps/cwp011616.pdf`.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*(2), 363–90.

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