Econometrica Supplementary Material

SUPPLEMENT TO "CONSTRAINED CONDITIONAL MOMENT RESTRICTION MODELS"

(*Econometrica*, Vol. 91, No. 2, March 2023, 709–736)

VICTOR CHERNOZHUKOV
Department of Economics, M.I.T.

WHITNEY K. NEWEY
Department of Economics, M.I.T.

ANDRES SANTOS
Department of Economics, U.C.L.A.

THIS SUPPLEMENT IS ORGANIZED AS FOLLOWS. Section A.1 provides a review of AM spaces. Section A.2 specializes the general results of Section 3 to three additional examples: (i) GMM, (ii) quantile treatment effects, and (iii) the Slutsky restriction in a partially linear model. The proofs for all results can be found in the working paper Chernozhukov, Newey, and Santos (2022).

A.1. AM SPACES

We provide a brief introduction to AM spaces and refer the reader to Chapters 8 and 9 of Aliprantis and Border (2006) for a more detailed exposition. Before proceeding, we first recall the definitions of a partially ordered set and a lattice.

DEFINITION A.1.1: A *partially ordered set* $(\mathbf{G}, \geq)$ is a set $\mathbf{G}$ with a partial order relationship $\geq$ defined on it—that is, $\geq$ is a transitive ($x \geq y$ and $y \geq z$ implies $x \geq z$), reflexive ($x \geq x$), and antisymmetric ($x \geq y$ implies the negation of $y \geq x$) relation.

DEFINITION A.1.2: A *lattice* is a partially ordered set $(\mathbf{G}, \geq)$ such that any pair $x, y \in \mathbf{G}$ has a least upper bound (denoted $x \vee y$) and a greatest lower bound (denoted $x \wedge y$).

Whenever $\mathbf{G}$ is both a vector space and a lattice, it is possible to define objects that depend on both the vector space and lattice operations. In particular, for $x \in \mathbf{G}$, we define the positive part $x^+ \equiv x \vee 0$, the negative part $x^- \equiv (-x) \vee 0$, and the absolute value $|x| \equiv x \vee (-x)$. It is also natural to demand that the order relation $\geq$ interact with the algebraic operations in a manner analogous to that of $\mathbf{R}$, that is, to have

$$x \geq y \quad \text{implies} \quad x + z \geq y + z \quad \text{for each } z \in \mathbf{G}, \tag{A.1}$$

$$x \geq y \quad \text{implies} \quad \alpha x \geq \alpha y \quad \text{for each } 0 \leq \alpha \in \mathbf{R}. \tag{A.2}$$

A complete normed vector space that shares these familiar properties of $\mathbf{R}$ under a given order relation $\geq$ is referred to as a *Banach lattice*. Formally, we define the following:

Victor Chernozhukov: vchern@mit.edu
Whitney K. Newey: wnewey@mit.edu
Andres Santos: andres@econ.ucla.edu

Whitney K. Newey's research supported by NSF Grant 1757140. Andres Santos' research supported by NSF Grant SES-1426882.

DEFINITION A.1.3: A Banach space $\mathbf{G}$ with norm $\|\cdot\|_{\mathbf{G}}$ is a *Banach lattice* if (i) $\mathbf{G}$ is a lattice under $\geq$, (ii) $\|x\|_{\mathbf{G}} \leq \|y\|_{\mathbf{G}}$ when $|x| \leq |y|$, (iii) (A.1) and (A.2) hold.

An AM space is a Banach lattice in which the maximum of the norms of any two positive elements is equal to the norm of the maximums of the two elements.

DEFINITION A.1.4: A Banach lattice $\mathbf{G}$ is called an AM space if, for any elements $0 \leq x, y \in \mathbf{G}$, it follows that $\|x \vee y\|_{\mathbf{G}} = \max\{\|x\|_{\mathbf{G}}, \|y\|_{\mathbf{G}}\}$.

In certain Banach lattices, there may exist an element $\mathbf{1}_{\mathbf{G}} > 0$ called an *order unit* such that, for any $x \in \mathbf{G}$, there exists a $0 < \lambda \in \mathbf{R}$ for which $|x| \leq \lambda \mathbf{1}_{\mathbf{G}}$; for example, in $\mathbf{R}^d$, the vector $(1, \dots, 1)'$ is an order unit. The order unit $\mathbf{1}_{\mathbf{G}}$ can be used to define

$$\|x\|_\infty \equiv \inf\{\lambda > 0 : |x| \leq \lambda \mathbf{1}_{\mathbf{G}}\}, \tag{A.3}$$

which is a norm on $\mathbf{G}$. In principle, $\|\cdot\|_\infty$ need not be related to the original norm $\|\cdot\|_{\mathbf{G}}$. However, if $\mathbf{G}$ is an AM space, then $\|\cdot\|_{\mathbf{G}}$ and $\|\cdot\|_\infty$ are equivalent in that they generate the same topology. Hence, we refer to $\mathbf{G}$ as an *AM space with unit* $\mathbf{1}_{\mathbf{G}}$ if: (i) $\mathbf{G}$ is an AM space, (ii) $\mathbf{1}_{\mathbf{G}}$ is an order unit in $\mathbf{G}$, and (iii) the norm of $\mathbf{G}$ equals $\|\cdot\|_\infty$.

A.2. ILLUSTRATIVE EXAMPLES

In this section, we examine special cases of our general analysis and illustrate both how to implement our procedure and how to verify the assumptions in the main text.

A.2.1. *Generalized Method of Moments*

Our first example concerns the generalized method of moments (GMM) model of Hansen (1982). We assume the parameter of interest θ_0 is identified as the unique solution to

$$E_P[\rho(X, \theta_0)] = 0, \tag{A.4}$$

where $X \in \mathbf{X}$ is distributed according to $P \in \mathbf{P}$ and $\rho : \mathbf{X} \times \Theta \to \mathbf{R}^{\mathcal{J}}$. This model maps into our general framework by letting $Z_\jmath = 1$ for all $1 \leq \jmath \leq \mathcal{J}$. Moreover, since we have assumed θ_0 is identified, the hypothesis testing problem simplifies to

$$H_0 : \theta_0 \in R, \qquad H_1 : \theta_0 \notin R.$$

The set R is, as in the main text, defined by equality and inequality restrictions. In particular, for known functions $\Upsilon_F : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_F}$ and $\Upsilon_G : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_G}$, we set

$$R \equiv \{\theta \in \mathbf{R}^{d_\theta} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}. \tag{A.5}$$

To verify Assumptions 3.1(ii),(iii), note $\mathbf{R}^d$ is a Banach space under any norm $\|\cdot\|_p$ with $1 \leq p \leq \infty$, so for concreteness we set $\mathbf{B} = \mathbf{R}^{d_\theta}$, $\mathbf{F} = \mathbf{R}^{d_F}$, and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{F}} = \|\cdot\|_2$. The space $\mathbf{R}^d$ is, in addition, a lattice under the standard pointwise partial order

$$a \leq b \quad \text{if and only if} \quad a_i \leq b_i \quad \text{for all } 1 \leq i \leq d \tag{A.6}$$

for any $(a_1, \dots, a_d)' = a$ and $(b_1, \dots, b_d)' = b$ in $\mathbf{R}^d$, while the least upper bound equals

$$a \vee b = (\max\{a_1, b_1\}, \dots, \max\{a_d, b_d\})'.$$

The vector $(1,\dots,1)'$ is an order unit in $\mathbf{R}^d$ under the partial order in (A.6). As discussed in Section A.1 of this supplement, the order unit induces the norm

$$\{\inf \lambda > 0 : |a| \le \lambda(1,\dots,1)'\} = \max_{1\le i\le d} |a_i|,$$

which corresponds to the usual $\|\cdot\|_\infty$ norm on $\mathbf{R}^d$. Hence, by setting $\mathbf{G} = \mathbf{R}^{d_G}$, $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$, and $\mathbf{1_G} = (1,\dots,1)'$, we verify the requirements of Assumption 3.1(ii),(iii).

Since the parameter space Θ is finite dimensional and all moment restrictions are unconditional, we may set $\Theta_n = \Theta$ and $k_n = \mathcal{J}$ for all n. We base our test statistic on quadratic forms in the moments ($p = 2$), which implies $Q_n(\theta)$ is given by

$$Q_n(\theta) \equiv \left\| \hat{\Sigma}_n \left\{ \frac{1}{n} \sum_{i=1}^n \rho(X_i, \theta) \right\} \right\|_2 .$$

In what follows, we consider tests based on both the un-centered statistic $I_n(R)$ and the re-centered statistic $I_n(R) - I_n(\Theta)$. To this end, we impose the following:

ASSUMPTION A.2.1: *(i)* $\{X_i\}_{i=1}^n$ *is i.i.d. with* $X_i \sim P \in \mathbf{P}$; *(ii) for each* $P \in \mathbf{P}_0$, *there exists a unique* $\theta_0 \in \Theta$ *solving* (A.4); *(iii)* Θ *is convex and compact.*

ASSUMPTION A.2.2: *(i) The function* $\rho(x,\cdot) : \Theta \to \mathbf{R}^{\mathcal{J}}$ *is twice differentiable for all* x; *(ii)* $E_P[\sup_{\theta\in\Theta} \|\rho(X,\theta)\|_2^3]$, $E_P[\sup_{\theta\in\Theta} \|\nabla_\theta \rho(X,\theta)\|_{o,2}^2]$, $E_P[\sup_{\theta\in\Theta} \|\nabla_\theta^2 \rho_J(X,\theta)\|_{o,2}^{1+\delta}]$ *are finite and bounded uniformly in* $P \in \mathbf{P}$ *for some* $\delta > 0$.

ASSUMPTION A.2.3: *(i)* $\inf_{P\in\mathbf{P}_0} \inf_{\theta\in\Theta:\|\theta-\theta_0\|_2\ge\epsilon} \|E_P[\rho(X,\theta)]\|_2 > 0$ *for all* $\epsilon > 0$; *(ii) the singular values of* $E_P[\nabla_\theta \rho(X,\theta_0)]$ *are bounded away from zero in* $P \in \mathbf{P}_0$.

ASSUMPTION A.2.4: *(i)* $\|\hat{\Sigma}_n - \Sigma_P\|_{o,2} = O_P(n^{-1/2})$ *uniformly in* $P \in \mathbf{P}$; *(ii)* Σ_P *is invertible and* $\|\Sigma_P\|_{o,2}$ *and* $\|\Sigma_P^{-1}\|_{o,2}$ *are bounded uniformly in* $P \in \mathbf{P}$.

In Assumption A.2.2, we focus on differentiable moments for simplicity. Assumption A.2.3 essentially imposes strong identification of θ_0 and hence guarantees that θ_0 can be consistently estimated uniformly in $P \in \mathbf{P}_0$—recall that θ_0 depends on P through (A.4), though the dependence is left implicit in the notation. Finally, Assumption A.2.4 states the requirements on the weighting matrix $\hat{\Sigma}_n$.

In what follows, we set the local parameter spaces $V_n(\theta, R|\ell)$ and $V_n(\theta, \Theta|\ell)$ to equal

$$V_n(\theta, R|\ell) = \{h \in \mathbf{R}^{d_\theta} : \theta + h/\sqrt{n} \in \Theta \cap R \text{ and } \|h/\sqrt{n}\|_2 \le \ell\},$$
$$V_n(\theta, \Theta|\ell) = \{h \in \mathbf{R}^{d_\theta} : \theta + h/\sqrt{n} \in \Theta \text{ and } \|h/\sqrt{n}\|_2 \le \ell\}.$$

Setting $\mathbb{D}_P(\theta_0)[h] \equiv E_P[\nabla_\theta \rho(X,\theta_0)]h$ and letting $\mathbb{W}_P(\theta_0) \sim N(0, \mathrm{Var}_P\{\rho(X,\theta_0)\})$, we then denote the variables to which $I_n(R)$ and $I_n(\Theta)$ will be coupled to by

$$U_P(R|\ell_n) \equiv \inf_{h\in V_n(\theta_0,R|\ell_n)} \|\mathbb{W}_P(\theta_0) + \mathbb{D}_P(\theta_0)[h]\|_{\Sigma_P,2},$$
$$U_P(\Theta|\ell_n) \equiv \inf_{h\in V_n(\theta_0,\Theta|\ell_n)} \|\mathbb{W}_P(\theta_0) + \mathbb{D}_P(\theta_0)[h]\|_{\Sigma_P,2}.$$

Our distributional approximations follow immediately from Theorem 3.1(ii).

THEOREM A.2.1: *Let Assumptions A.2.1, A.2.2, A.2.3, and A.2.4 hold*, Υ_F *and* Υ_G *be continuous, and set* $a_n = \sqrt{\log(n)}/n^{1/(10+5d_\theta)}$. *Then, for any* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $(\ell_n \vee \ell_n^{\mathrm{u}})\sqrt{\log(1/\ell_n \vee \ell_n^{\mathrm{u}})} = o(a_n)$ *and* $n^{-1/2} = o(\ell_n \vee \ell_n^{\mathrm{u}})$, *we have, uniformly in* $P \in \mathbf{P}_0$,

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n),$$

$$I_n(R) - I_n(\Theta) = U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

The rate of coupling $a_n = \sqrt{\log(n)}/n^{1/(10+5d_\theta)}$ obtained in Theorem A.2.1 suffices for both the empirical process and bootstrap coupling. While the rate is adequate for our purposes, it can be improved under additional moment restrictions. Here, we rely on Yurinskii (1977) both to illustrate the diversity of coupling arguments that can be employed to verify Assumption 3.3(i) and to impose only the weak third moment restriction of Assumption A.2.2(ii).

Our next goal is to obtain bootstrap approximations to the distributions of $U_P(R|\ell_n)$ and $U_P(\Theta|\ell_n^{\mathrm{u}})$. To this end, we write $\Upsilon_F(\theta) = (\Upsilon_{F,1}(\theta), \ldots, \Upsilon_{F,d_F}(\theta))'$ and $\Upsilon_G(\theta) = (\Upsilon_{G,1}(\theta), \ldots, \Upsilon_{G,d_G}(\theta))'$, for any $\epsilon > 0$ we define $B^\epsilon \equiv \bigcup_{P\in\mathbf{P}_0}\{\theta : \|\theta - \theta_0\|_2 \leq \epsilon\}$ (where recall θ_0 implicitly depends on P through (A.4)), and impose the following:

ASSUMPTION A.2.5: *For some* $\epsilon > 0$: (*i*) $B^\epsilon \subseteq \Theta$; (*ii*) Υ_F *and* Υ_G *are twice differentiable on* B^ϵ; (*iii*) $\|\nabla\Upsilon_F(\theta)\|_{o,2}$ *and* $\|\nabla\Upsilon_G(\theta)\|_{o,2}$ *are bounded on* B^ϵ; (*iv*) $\|\nabla^2\Upsilon_{F,j}(\theta)\|_{o,2}$ *is bounded on* B^ϵ *for* $1 \leq j \leq d_F$; (*v*) $\|\nabla^2\Upsilon_{G,j}(\theta)\|_{o,2}$ *is bounded on* B^ϵ *for* $1 \leq j \leq d_G$; (*vi*) $\nabla\Upsilon_F(\theta)$ *has full row-rank on* B^ϵ.

ASSUMPTION A.2.6: *Either* (*i*) $\Upsilon_F : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_F}$ *is affine, or* (*ii*) *there is an* $\epsilon > 0$ *and* $M < \infty$ *such that the singular values of* $\nabla\Upsilon_F(\theta)'$ *are bounded away from zero uniformly in* $\theta \in B^\epsilon$, *and for every* $P \in \mathbf{P}_0$ *there is an* $h \in \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ *with* $\|h\|_2 \leq M$ *satisfying* $\Upsilon_{G,j}(\theta_0) + \nabla\Upsilon_{G,j}(\theta_0)[h] \leq -\epsilon$ *for all* $1 \leq j \leq d_G$.

In order to describe our bootstrap procedure in this application, we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote the minimizers of Q_n over $\Theta \cap R$ and Θ, respectively. Employing $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$, we obtain estimators for the distribution of $\mathbb{W}_P(\theta_0)$ and for $\mathbb{D}_P(\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i\left\{\rho(X_i, \theta) - \frac{1}{n}\sum_{j=1}^n \rho(X_j, \theta)\right\}, \tag{A.7}$$

$$\hat{\mathbb{D}}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \nabla_\theta\rho(X_i, \theta), \tag{A.8}$$

at $\theta = \hat{\theta}_n$ and $\theta = \hat{\theta}_n^{\mathrm{u}}$, where recall $\{\omega_i\}_{i=1}^n$ is an i.i.d. sample independent of $\{X_i\}_{i=1}^n$ with $\omega_i \sim N(0, 1)$. We note that because moments are differentiable, we employ an analytical derivative in (A.8) instead of the numerical derivative studied in Section 3.

With regard to the local parameter space, we note that the construction of $\hat{V}_n(\theta, R|\ell)$ requires the bound K_g on the second derivative of Υ_G (as specified in Assumption 3.8). In particular, Assumption A.2.5(v) implies Assumption 3.8 is satisfied with

$$K_g \equiv \max_{1\leq j\leq d_G}\sup_{\theta\in B^\epsilon}\left\|\nabla_\theta^2\Upsilon_{G,j}(\theta)\right\|_{o,2}.$$

If an a priori bound on the second derivative is not available, then it is also possible to simply substitute K_g with the data driven choice

$$\hat{K}_g \equiv \max_{1\leq j\leq d_G} \sup_{\theta\in\Theta:\|\theta-\hat{\theta}_n\|_2\leq r_n} \left\|\nabla_\theta^2 \Upsilon_{G,j}(\theta)\right\|_{o,2},$$

where we discuss the choice of r_n below. Given K_g (or $\hat{K}_g$), we set $G_n(\theta)$ to equal

$$G_n(\theta) = \left\{h \in \mathbf{R}^{d_\theta} : \Upsilon_{G,j}\left(\theta + \frac{h}{\sqrt{n}}\right) \leq \max\left\{\Upsilon_{G,j}(\theta) - K_g r_n\left\|\frac{h}{\sqrt{n}}\right\|_2, -r_n\right\} \text{ for all } j\right\}.$$

In this application we may additionally specify ℓ_n to be infinite, and hence we set

$$\hat{V}_n(\theta, R|+\infty) = \left\{h \in \mathbf{R}^{d_\theta} : h \in G_n(\theta) \text{ and } \Upsilon_F\left(\theta + \frac{h}{\sqrt{n}}\right) = 0\right\}.$$

The approximations to the distributions of $I_n(R)$ and $I_n(\Theta)$ are then given by the laws of $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(\Theta|+\infty)$ conditional on the data, where

$$\hat{U}_n(R|+\infty) \equiv \inf_{h\in\hat{V}_n(\hat{\theta}_n,R|+\infty)} \left\|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\right\|_{\hat{\Sigma}_n,2},$$

$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{h\in\mathbf{R}^{d_\theta}} \left\|\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n(\hat{\theta}_n^{\mathrm{u}})[h]\right\|_{\hat{\Sigma}_n,2}.$$

The validity of these distributional approximations follows from Theorem 3.2.

THEOREM A.2.2: *Let Assumptions A.2.1, A.2.2, A.2.3, A.2.4, A.2.5, and A.2.6 hold, set* $a_n = \sqrt{\log(n)}/n^{1/(10+5d_\theta)}$, *and let* $n^{-1/2} = o(r_n)$. *Then: there are sequences* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $(\ell_n \vee \ell_n^{\mathrm{u}})^2\sqrt{\log(1/(\ell_n \vee \ell_n^{\mathrm{u}}))} = o(a_n n^{-1/2})$, $\ell_n = o(r_n)$, *and* $n^{-1/2} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$ *for which it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$\hat{U}_n(R|+\infty) \geq U_P^\star(R|\ell_n) + o_P(a_n),$$

$$\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty) \geq U_P^\star(R|\ell_n) - U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

Crucially, note that any sequences ℓ_n and ℓ_n^{u} satisfying the conditions of Theorem A.2.2 also satisfy the conditions of Theorem A.2.1. Therefore, Theorems A.2.2 and A.2.1 together establish the validity of employing the laws of $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(\Theta|+\infty)$ conditional on the data to approximate the laws of $I_n(R)$ and $I_n(\Theta)$. In particular, for a level α test, we may compare the test statistic $I_n(R)$ to the critical value

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|+\infty)) \equiv \inf\{c : P(\hat{U}_n(R|+\infty) \leq c|\{X_i\}_{i=1}^n) \geq 1-\alpha\}.$$

Similarly, for the re-centered statistic $I_n(R) - I_n(\Theta)$, valid critical values are given by

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty)) \\ \equiv \inf\{c : P(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty) \leq c|\{X_i\}_{i=1}^n) \geq 1-\alpha\}.$$

These approximations are valid under the requirement that r_n satisfy $r_n\sqrt{n} \to \infty$. Intuitively, the bandwidth r_n is meant to reflect a bound on the distance between $\hat{\theta}_n$ and θ_0.

For a data driven choice of r_n, we may therefore employ a bootstrap estimate of an upper quantile of the distribution of the *unconstrained* estimator. Specifically, for $\hat{\theta}_n^{u\star}$ the bootstrapped version of $\hat{\theta}_n^u$, we may set $\hat{r}_n$ to be given by

$$\hat{r}_n \equiv \inf\{c : P(\|\hat{\theta}_n^{u\star} - \hat{\theta}_n^u\|_2 \leq c|\{X_i\}_{i=1}^n) \geq 1 - \gamma_n\}$$

for $\gamma_n \to 0$ as the sample size n tends to infinity, and employ $\hat{r}_n$ in place of r_n.

A.2.2. *Consumer Demand*

We base our next example on a longstanding literature aiming to replace parametric assumptions with shape restrictions implied by economic theory (Matzkin (1994)). Specifically, suppose that quantity demanded by individual i, denoted Q_i, satisfies

$$Q_i = g_0(S_i, Y_i) + W_i'\gamma_0 + U_i,$$

where $S_i \in \mathbf{R}_+$ denotes price, $Y_i \in \mathbf{R}_+$ denotes income, and $W_i \in \mathbf{R}^{d_w}$ is a set of covariates. In addition, we assume there is an instrument Z_i yielding the restriction

$$E_P[Q - g_0(S, Y) - W'\gamma_0|Z] = 0. \tag{A.9}$$

For instance, under exogeneity of prices, we may let $Z = (S, Y, W')'$ as in Blundell, Horowitz, and Parey (2012). Alternatively, if there is a concern that prices are endogenous, then we may set $Z = (I, Y, W')'$ for I an instrument for S, as in Blundell, Horowitz, and Parey (2017).

Our goal is to conduct inference on the level of demand at a particular price-income pair (s_0, y_0) while imposing that the function g_0 satisfies the Slutsky restriction

$$\frac{\partial}{\partial s} g_0(s, y) + g_0(s, y)\frac{\partial}{\partial y} g_0(s, y) \leq 0. \tag{A.10}$$

To map this problem into our framework, we assume that for some set Ω, $(S, Y) \in \Omega \subseteq \mathbf{R}_+^2$ with probability 1 for all $P \in \mathbf{P}$ and impose that $g_0 \in C_B^1(\Omega)$, where

$$C_B^m(\Omega) \equiv \{g : \Omega \to \mathbf{R} \text{ s.t. } \|g\|_{m,\infty} < \infty\}, \qquad \|g\|_{m,\infty} \equiv \sup_{0\leq\alpha\leq m} \sup_{(s,y)\in\Omega} |\nabla^\alpha g(s, y)|.$$

Since $\theta_0 \equiv (g_0, \gamma_0)$ with $\gamma_0 \in \mathbf{R}^{d_w}$, we set $\mathbf{B} = C_B^1(\Omega) \times \mathbf{R}^{d_w}$ and for any $(g, \gamma) = \theta \in \mathbf{B}$ let $\|\theta\|_{\mathbf{B}} = \max\{\|g\|_{1,\infty}, \|\gamma\|_2\}$. We also note that $X = (Q, S, Y, W)$ and

$$\rho(X, \theta) = Q - g(S, Y) - W'\gamma. \tag{A.11}$$

We will assume $\theta_0 \equiv (g_0, \gamma_0)$ is identified by (A.9). Hence, we can think of θ_0 as a function of P through (A.9), though we leave such dependence implicit in the notation.

In order to impose the Slutsky restriction in (A.10), we let $\mathbf{G} = C_B^0(\Omega)$ and $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$, where with some abuse of notation we write $\|\cdot\|_\infty$ in place of $\|\cdot\|_{0,\infty}$. The space $C_B^0(\Omega)$ is a Banach lattice under the standard pointwise ordering given by

$$a \leq b \quad \text{if and only if} \quad a(s, y) \leq b(s, y) \quad \text{for all } (s, y) \in \Omega \tag{A.12}$$

for any $a, b \in C^0_B(\Omega)$. The constant function $\mathbf{c} \in C^0_B(\Omega)$ satisfying $\mathbf{c}(s, y) = 1$ for all $(s, y) \in \Omega$ is an order unit under the partial ordering in (A.12). Its induced norm is

$$\{\inf \lambda > 0 : |a| \leq \lambda \mathbf{c}\} = \sup_{(s,y)\in\Omega} |a(s, y)|,$$

which coincides with the norm $\|\cdot\|_\infty$ on $C^0_B(\Omega)$, and we therefore set $\mathbf{1_G} = \mathbf{c}$. To encode the Slutsky restriction in (A.10), we then let the map $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ equal

$$\Upsilon_G(\theta)(s, y) = \frac{\partial}{\partial s} g(s, y) + g(s, y) \frac{\partial}{\partial y} g(s, y) \tag{A.13}$$

for any $\theta = (g, \gamma) \in \mathbf{B}$. Finally, to test whether the level of demand at a prescribed price s_0 and income y_0 equals a hypothesized value c_0, we set $\mathbf{F} = \mathbf{R}$, $\|\cdot\|_\mathbf{F} = |\cdot|$, and

$$\Upsilon_F(\theta) = g(s_0, y_0) - c_0 \tag{A.14}$$

for any $\theta = (g, \gamma) \in \mathbf{B}$. By setting $R = \{\theta \in \mathbf{B} : \Upsilon_G(\theta) \leq 0 \text{ and } \Upsilon_F(\theta) = 0\}$ and conducting test inversion (over different values of c_0) of the null hypothesis

$$H_0 : \theta_0 \in R, \qquad H_1 : \theta_0 \notin R,$$

we may obtain a confidence region for the level of demand at price s_0 and income y_0.

We set the parameter space to be a ball in $\mathbf{B}$ under $\|\cdot\|_\mathbf{B}$ by letting Θ be equal to

$$\Theta \equiv \{(g, \gamma) \in C^1_B(\Omega) \times \mathbf{R}^{d_w} : \|g\|_{1,\infty} \leq C_0 \text{ and } \|\gamma\|_2 \leq C_0\} \tag{A.15}$$

for some $C_0 < \infty$. Given a sequence of approximating functions $\{p_j\}_{j=1}^{j_n}$, we then let $p^{j_n}(s, y) \equiv (p_1(s, y), \dots, p_{j_n}(s, y))'$ and set the sieve Θ_n to equal

$$\Theta_n \equiv \{(p^{j_n\prime}\beta, \gamma) : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0 \text{ and } \|\gamma\|_2 \leq C_0\}.$$

Similarly, for a sequence $\{q_k\}_{k=1}^{k_n}$ of transformations of the conditioning variable Z, we let $q^{k_n}(z) \equiv (q_1(z), \dots, q_{k_n}(z))'$. We base our test statistic on the quadratic forms

$$Q_n(\theta) \equiv \left\| \hat{\Sigma}_n \left\{ \frac{1}{n} \sum_{i=1}^n (Q_i - g(S_i, Y_i) - W_i'\gamma) q^{k_n}(Z_i) \right\} \right\|_2$$

for some $k_n \times k_n$ weighting matrix $\hat{\Sigma}_n$ and every $(g, \gamma) = \theta \in \Theta$. The statistics $I_n(R)$ and $I_n(\Theta)$ simply equal the minimums of $\sqrt{n} Q_n(\theta)$ over $\Theta_n \cap R$ and Θ_n, respectively.

The next assumptions suffice for obtaining a strong approximation. In their statement, the notation $\underline{\text{sing}}\{A\}$ denotes the smallest singular value of a matrix A.

ASSUMPTION A.2.7: (*i*) *$\{X_i, Z_i\}_{i=1}^n$ is i.i.d. with (X, Z) distributed according to $P \in \mathbf{P}$*; (*ii*) *for Θ as in* (A.15) *and each $P \in \mathbf{P}_0$, there exists a unique $\theta_0 \in \Theta$ satisfying $E_P[\rho(X, \theta_0)|Z] = 0$*; (*iii*) *the support of (Q, W) is bounded uniformly in $P \in \mathbf{P}$.*

ASSUMPTION A.2.8: (*i*) $\sup_{(s,y)} \|p^{j_n}(s, y)\|_2 \lesssim \sqrt{j_n}$; (*ii*) $\sup_{(s,y)} \|\partial_a p^{j_n}(s, y)\|_2 \lesssim j_n^{3/2}$ *for $a \in \{s, y\}$*; (*iii*) *the eigenvalues of $E_P[p^{j_n}(S, Y) p^{j_n}(S, Y)']$ are bounded away from zero and infinity uniformly in $P \in \mathbf{P}$ and j_n*; (*iv*) *for each $P \in \mathbf{P}_0$, there is a $\Pi_n \theta_0 = (g_n, \gamma_0) \in \Theta_n \cap R$ with* $\sup_{P \in \mathbf{P}_0} \|E_P[(g_0(S, Y) - g_n(S, Y)) q^{k_n}(Z)\|_2 = o((n \log(n))^{-1/2})$.

ASSUMPTION A.2.9: *(i)* $\max_{1\le k\le k_n}\|q_k\|_\infty \lesssim \sqrt{k_n}$; *(ii)* $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ *has eigenvalues bounded uniformly in* $P\in\mathbf{P}$ *and* k_n; *(iii)* $\mathrm{s}_n \equiv \inf_{P\in\mathbf{P}} \underline{\mathrm{sing}}\{E_P[q^{k_n}(Z)(p^{j_n}(S,Y)',W')]\}$ *satisfies* $0<\mathrm{s}_n = O(1)$; *(iv)* $j_n^2k_n^3\log^3(n)=o(n)$ *and* $k_n^2 j_n\log^{3/2}(1+k_n)/(\mathrm{s}_n\sqrt{n})(1\vee\sqrt{\log(\mathrm{s}_n\sqrt{n}/k_n)})=o((\log(n))^{-1/2})$.

ASSUMPTION A.2.10: *(i)* $\|\hat{\Sigma}_n-\Sigma_P\|_{o,2}=o_P((k_n\sqrt{j_n}\log^{3/2}(n))^{-1})$ *uniformly in* $P\in\mathbf{P}$; *(ii)* Σ_P *is invertible and* $\|\Sigma_P\|_{o,2}$ *and* $\|\Sigma_P^{-1}\|_{o,2}$ *are bounded in* $P\in\mathbf{P}$ *and* k_n.

Assumption A.2.7(iii) requires (Q,W) to be bounded, which enables us to apply the recent coupling results by Zhai (2018). Alternatively, Assumption A.2.7(iii) can be relaxed under appropriate tail conditions. Assumptions A.2.8(i)–(iii) are standard requirements on Θ_n that can be satisfied by, for example, tensor product wavelets or B-splines (Newey (1997), Chen (2007), Belloni, Chernozhukov, Chetverikov, and Kato (2015), Chen and Christensen (2018)). Assumption A.2.8(iv) pertains to the approximating requirements on the sieve; see Remarks A.2.1 and A.2.2 below. In turn, Assumptions A.2.9(i),(ii) impose standard requirements on $\{q_k\}_{k=1}^{k_n}$. Assumptions A.2.9(iii),(iv) contain the required rate conditions, which are governed by s_n—a parameter that is proportional to ν_n^{-1} (as in Assumption 3.4) and is closely linked to the degree of ill-posedness; see Remark A.2.2 below. Finally, Assumption A.2.10 states the conditions on the weighting matrix $\hat{\Sigma}_n$.

In this application, we may set $\|\theta\|_{\mathbf{E}}=\sup_{P\in\mathbf{P}}\|g\|_{P,2}+\|\gamma\|_2$ for any $(g,\gamma)\in\Theta$. Since, in addition, any $\theta=(g,\gamma)\in\Theta_n\cap R$ has the structure $g=p^{j_n\prime}\beta$, we have

$$V_n(\theta,R|\ell)=\biggl\{(p^{j_n\prime}\beta_h,\gamma_h):\Bigl\|g+\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\Bigr\|_{1,\infty}\le C_0 \text{ and } \Bigl\|\gamma+\frac{\gamma_h}{\sqrt{n}}\Bigr\|_2\le C_0, \tag{A.16}$$

$$p^{j_n}(s_0,y_0)'\beta_h=0, \tag{A.17}$$

$$\frac{\partial}{\partial s}\Bigl(g+\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\Bigr)+\Bigl(g+\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\Bigr)\frac{\partial}{\partial y}\Bigl(g+\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\Bigr)\le 0, \tag{A.18}$$

$$\sup_{P\in\mathbf{P}}\bigl\|p^{j_n\prime}\beta_h\bigr\|_{P,2}+\|\gamma_h\|_2\le \ell\sqrt{n}\biggr\}, \tag{A.19}$$

where constraint (A.16) corresponds to $(\theta+h/\sqrt{n})\in\Theta_n$, constraints (A.17) and (A.18) impose $\theta+h/\sqrt{n}\in R$, and constraint (A.19) imposes $\|h/\sqrt{n}\|_{\mathbf{E}}\le\ell$. Similarly,

$$V_n(\theta,\Theta|\ell)=\biggl\{(p^{j_n\prime}\beta_h,\gamma_h):\Bigl\|g+\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\Bigr\|_{1,\infty}\le C_0 \text{ and } \Bigl\|\gamma+\frac{\gamma_h}{\sqrt{n}}\Bigr\|_2\le C_0, \tag{A.20}$$

$$\sup_{P\in\mathbf{P}}\bigl\|p^{j_n\prime}\beta_h\bigr\|_{P,2}+\|\gamma_h\|_2\le \ell\sqrt{n}\biggr\}. \tag{A.21}$$

Finally, recall that $\mathbb{W}_P(\theta)\sim N(0,\mathrm{Var}_P\{\rho(X,\theta)q^{k_n}(Z)\})$ and define $\mathbb{D}_P$ to be given by

$$\mathbb{D}_P[h]\equiv -E_P\bigl[q^{k_n}(Z)\bigl(p^{j_n}(S,Y)'\beta_h+W'\gamma_h\bigr)\bigr]$$

for any $h=(p^{j_n\prime}\beta_h,\gamma_h)$. Given these definitions, note that, for any ℓ_n, we have that

$$U_P(R|\ell_n)\equiv \inf_{h\in V_n(\Pi_n\theta_0,R|\ell_n)}\bigl\|\mathbb{W}_P(\Pi_n\theta_0)+\mathbb{D}_P[h]\bigr\|_{\Sigma_P,2},$$

$$U_P(\Theta|\ell_n)\equiv \inf_{h\in V_n(\Pi_n\theta_0,\Theta|\ell_n)}\bigl\|\mathbb{W}_P(\Pi_n\theta_0)+\mathbb{D}_P[h]\bigr\|_{\Sigma_P,2}.$$

Theorem 3.1(ii) immediately yields the following distributional approximations.

THEOREM A.2.3: *Let Assumptions A.2.7–A.2.10 hold, and* $a_n = (\log(n))^{-1/2}$. *Then, for any* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $k_n\sqrt{j_n \log(1+k_n)}(\ell_n \vee \ell_n^{\mathrm{u}})\sqrt{\log(\sqrt{j_n}/(\ell_n \vee \ell_n^{\mathrm{u}}))} = o(a_n)$ *and* $k_n\sqrt{j_n}\log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$, *it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n),$$
$$I_n(R) - I_n(\Theta) = U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

To obtain bootstrap estimates of the distributional approximations in Theorem A.2.3, we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote the minimizers of Q_n over $\Theta_n \cap R$ and Θ_n, respectively. For $\rho(\cdot, \theta)$ as in (A.11), we approximate the law of $\mathbb{W}_P(\Pi_n\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i\left\{q^{k_n}(Z_i)\rho(X_i, \theta) - \frac{1}{n}\sum_{j=1}^{n}q^{k_n}(Z_j)\rho(X_j, \theta)\right\},$$

at $\theta = \hat{\theta}_n$ and $\theta = \hat{\theta}_n^{\mathrm{u}}$, where $\{\omega_i\}_{i=1}^n$ is an i.i.d. sample independent of the data satisfying $\omega_i \sim N(0, 1)$. As our estimator for $\mathbb{D}_P[h]$, for any $h = (p^{j_n\prime}\beta_h, \gamma_h)$, we let

$$\hat{\mathbb{D}}_n[h] = -\frac{1}{n}\sum_{i=1}^{n}q^{k_n}(Z_i)\big(p^{j_n}(S_i, Y_i)'\beta_h + W_i'\gamma_h\big).$$

With regard to the local parameter space, we note that, in this application, Assumptions 3.8(i),(ii) are satisfied with $K_g = 2$. Therefore, we have

$$\begin{aligned}
&G_n(\hat{\theta}_n)\\
&\quad= \bigg\{h : \frac{\partial}{\partial s}p^{j_n}(s, y)'\bigg(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}}\bigg) + p^{j_n}(s, y)'\bigg(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}}\bigg)\frac{\partial}{\partial y}p^{j_n}(s, y)'\bigg(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}}\bigg)\\
&\quad\le \max\bigg\{\frac{\partial}{\partial s}p^{j_n}(s, y)'\hat{\beta}_n + p^{j_n}(s, y)'\hat{\beta}_n\frac{\partial}{\partial y}p^{j_n}(s, y)'\hat{\beta}_n - 2r_n\bigg\|\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\bigg\|_{1,\infty}, -r_n\bigg\}\bigg\}. \qquad \text{(A.22)}
\end{aligned}$$

Moreover, because $\rho(X, \cdot)$ and Υ_F are linear, we may set $\ell_n = +\infty$ and obtain that

$$\hat{V}_n(\hat{\theta}_n, R|+\infty) = \big\{h = \big(p^{j_n\prime}\beta_h, \gamma_h\big) : h \in G_n(\hat{\theta}_n) \text{ and } p^{j_n}(s_0, y_0)'\beta_h = 0\big\}.$$

Given the introduced notation, we define the statistics $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(\Theta|+\infty)$ by

$$\hat{U}_n(R|+\infty) \equiv \inf_{h\in\hat{V}_n(\hat{\theta}_n, R|+\infty)}\big\|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n[h]\big\|_{\hat{\Sigma}_n,2},$$
$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{h=(p^{j_n\prime}\beta_h, \gamma_h)}\big\|\hat{\mathbb{W}}_n\big(\hat{\theta}_n^{\mathrm{u}}\big) + \hat{\mathbb{D}}_n[h]\big\|_{\hat{\Sigma}_n,2}.$$

We impose one final assumption to establish the validity of the bootstrap. In the requirements below, it is helpful to recall θ_0 is implicitly a function of P through (A.9).

ASSUMPTION A.2.11: (*i*) *There is an* $\epsilon > 0$ *such that* $\|g_0\|_{1,\infty} \vee \|\gamma_0\|_2 \leq C_0 - \epsilon$ *for all* $P \in \mathbf{P}_0$; (*ii*) $\Pi_n\theta_0 = (g_n, \gamma_0) \in \Theta_n \cap R$ *satisfies* $\|g_n - g_0\|_{1,\infty} = o(1)$ *uniformly in* $P \in \mathbf{P}_0$; (*iii*) *the sequence* $r_n \downarrow 0$ *satisfies* $k_n j_n^2\sqrt{\log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(r_n/\sqrt{\log(n)})$; (*iv*) $k_n j_n^{3/4}(\mathcal{E}_n \vee \sqrt{\log(k_n)})\log^{1/4}(1+k_n) = o(n^{1/4}/\sqrt{\log(n)})$, *where* $\mathcal{E}_n \equiv \int_0^\infty \sqrt{\log(\epsilon, \mathcal{C}_n, \|\cdot\|_2)}\, d\epsilon$ *and* $\mathcal{C}_n \equiv \{\beta : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0\}$.

Assumptions A.2.11(i),(ii) suffice for verifying Assumption 3.12(ii). These requirements may be dropped at the expense of modifying $\hat{V}_n(\hat{\theta}_n, R| + \infty)$ to reflect the possible impact of $\Pi_n\theta_0$ being "near" the boundary of Θ_n. Assumption A.2.11(iii) imposes the rate conditions on r_n. Finally, Assumption A.2.11(iv) controls the "size" of the set of coefficients β corresponding to elements $p^{j_n\prime}\beta \in \Theta_n$ and suffices for verifying the bootstrap coupling requirement of Assumption 3.11. For instance, $\mathcal{E}_n \asymp j_n^{1/4}$ for tensor product B-splines, which implies a sufficient condition for Assumption A.2.11(iv) is that $k_n^4 j_n^4 \log^4(k_n) = o(n/\log^2(n))$. The rate requirements for a bootstrap coupling can be weakened if the test statistic is based on the $\|\cdot\|_\infty$-norm or under additional smoothness assumptions.

Our next result characterizes the properties of the proposed bootstrap statistics.

THEOREM A.2.4: *Let Assumptions* A.2.7, A.2.8, A.2.9, A.2.10, A.2.11 *hold, and* $a_n = (\log(n))^{-1/2}$. *Then, there are sequences* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $k_n j_n^2 \log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$, $\ell_n = o(r_n)$, *and* $k_n\sqrt{j_n \log(1+k_n)}(\ell_n \vee \ell_n^{\mathrm{u}})\sqrt{\log(\sqrt{j_n}/(\ell_n \vee \ell_n^{\mathrm{u}}))} = o(a_n)$, *for which it follows that, uniformly in* $P \in \mathbf{P}_0$, *we have*

$$\hat{U}_n(R| + \infty) \geq U_P^\star(R|\ell_n) + o_P(a_n),$$

$$\hat{U}_n(R| + \infty) - \hat{U}_n(\Theta| + \infty) \geq U_P^\star(R|\ell_n) - U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

Importantly, any sequences ℓ_n and ℓ_n^{u} satisfying the requirements of Theorem A.2.4 also satisfy the requirements of Theorem A.2.3. Hence, we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R| + \infty)) \equiv \inf\{c : P(\hat{U}_n(R| + \infty) \leq c|\{V_i\}_{i=1}^n) \geq 1 - \alpha\}$$

as a critical value for $I_n(R)$. Similarly, for the statistic $I_n(R) - I_n(\Theta)$, we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R| + \infty) - \hat{U}_n(\Theta| + \infty)) \equiv \inf\{c : P(\hat{U}_n(R| + \infty) - \hat{U}_n(\Theta| + \infty) \leq c|\{V_i\}_{i=1}^n) \geq 1 - \alpha\}.$$

REMARK A.2.1: Suppose for notational simplicity that there are no covariates W and let the marginal distribution of (S, Y, Z) be constant in $P \in \mathbf{P}$. If $Z = (S, Y)$ (i.e., (S, Y) is exogenous), we may set $q^{k_n}(Z) = p^{k_n}(S, Y)$ for some $k_n \geq j_n$. The singular value s_n can then be assumed to be bounded away from zero, and a sufficient condition for Assumption A.2.9(iv) is that $k_n^4 j_n^2 \log^5(n) = o(n)$. In order to appreciate the content of Assumption A.2.8(iv), suppose $\{p_j\}_{j=1}^\infty$ is an orthonormal basis such that

$$g_0 = \sum_{j=1}^\infty \beta_j p_j \quad \text{with } |\beta_j| = O(j^{-\gamma_\beta}).$$

Setting $\Pi_n^{\mathrm{u}} g_0 = \sum_{j=1}^{j_n} p_j \beta_j$, we obtain from a standard integral bound for a sum that

$$\left\| E_P\left[\left(g_0(S, Y) - \Pi_n^{\mathrm{u}} g_0(S, Y)\right) q^{k_n}(Z)\right]\right\|_2^2 \lesssim \sum_{j=j_n+1}^{k_n} \frac{1}{j^{2\gamma_\beta}} \lesssim \frac{1}{j_n^{2\gamma_\beta - 1}} - \frac{1}{k_n^{2\gamma_\beta - 1}}. \tag{A.23}$$

For instance, if $k_n - j_n = O(1)$, then the bound in (A.23) is of order $1/j_n^{2\gamma_\beta}$. Hence, provided the approximation errors by $\Pi_n^{\mathrm{u}} g_0$ and g_n (as in Assumption A.2.8(iv)) are of the same order when $g_0 \in R$, we obtain that Assumption A.2.8(iv) is equivalent to $\sqrt{n \log(n)}/j_n^{\gamma_\beta} = o(1)$ when $k_n - j_n = O(1)$. This approximation requirement is compatible with the condition $k_n^4 j_n^2 \log^5(n) = o(n)$ provided $\gamma_\beta > 3$.

REMARK A.2.2: Building on Remark A.2.1, suppose again there are no covariates W and the marginal distribution of (S, Y, Z) is constant in $P \in \mathbf{P}$, but now let (S, Y) be endogenous. A standard benchmark for nonparametric models with endogeneity is to assume the operator $g \mapsto E_P[g(S, Y)|Z]$ is compact, in which case there are orthonormal sequences of functions $\{\phi_j\}_{j=1}^{\infty}$ of (S, Y) and $\{\psi_j\}_{j=1}^{\infty}$ of Z satisfying

$$E_P\left[\phi_j(S, Y)|Z\right] = \lambda_j \psi_j(Z), \qquad E_P\left[\psi_j(Z)|S, Y\right] = \lambda_j \phi_j(S, Y),$$

where $\lambda_j > 0$ tends to zero. In addition suppose g_0 admits for an expansion satisfying

$$g_0 = \sum_{j=1}^{\infty} \beta_j \phi_j \quad \text{with } |\beta_j| = O\left(j^{-\gamma_\beta}\right),$$

and let $p^{j_n} = (\phi_1, \ldots, \phi_{j_n})'$, $q^{k_n} = (\psi_1, \ldots, \psi_{k_n})'$ with $k_n \geq j_n$ and $k_n - j_n = O(1)$, and set $\Pi_n^{\mathrm{u}} g_0 = \sum_{j=1}^{j_n} \phi_j \beta_j$. Provided the approximation errors of $\Pi_n^{\mathrm{u}} g_0$ and g_n (as in Assumption A.2.8(iv)) are of the same order when $g_0 \in R$, we then obtain

$$\left\| E_P\left[\left(g_0(S, Y) - g_n(S, Y)\right) q^{k_n}(Z)\right]\right\|_2 \lesssim \frac{\lambda_{j_n}}{j_n^{\gamma_\beta}}.$$

Moreover, direct calculation shows s_n, which is proportional to ν_n^{-1} as in Assumption 3.4, satisfies $\mathrm{s}_n = \lambda_{j_n}$ and hence equals the reciprocal of the sieve measure of ill-posedness (Blundell, Chen, and Kristensen (2007)). It follows that if $\lambda_j \asymp j^{-\gamma_\lambda}$, and $\gamma_\beta > 3$, then Assumptions A.2.8(iv) and A.2.9(iv) can be satisfied by setting $j_n \asymp n^{\kappa}$ with $(\gamma_\lambda + \gamma_\beta)^{-1} < 2\kappa < (3 + \gamma_\lambda)^{-1}$ and $k_n - j_n = O(1)$. Alternatively, if $\lambda_j = \exp\{-\gamma_\lambda j\}$, then Assumptions A.2.8(iv) and A.2.9(iv) can be satisfied when $\gamma_\beta > 4$ by setting, for example, $j_n = (\log(n) - \kappa \log(\log(n)))/2\gamma_\lambda$ with $7 < \kappa < 2\gamma_\beta - 1$ and $k_n - j_n = O(1)$.

A.2.3. *Quantile Treatment Effects*

For our next example, we study a nonparametric quantile treatment effect (QTE) model. Specifically, for an outcome $Y \in \mathbf{R}$, treatment $D \in [0, 1]$, instrument $Z \in \mathbf{R}$, and quantile $\tau \in (0, 1)$, we assume the parameter of interest θ_0 satisfies

$$P\left(Y \leq \theta_0(D)|Z\right) = \tau. \tag{A.24}$$

If D is randomly assigned, then we may set $D = Z$ and interpret $\nabla\theta_0$ as the τth quantile treatment effect (QTE). Alternatively, if $D \neq Z$, then we obtain the QTE model of Chernozhukov and Hansen (2005). To map (A.24) into our framework, we set

$$\rho(X,\theta) = 1\{Y \leq \theta(D)\} - \tau, \tag{A.25}$$

where $X = (Y, D) \in \mathbf{X} \equiv \mathbf{R} \times [0,1]$. In order to illustrate our conditions in a number of different settings, we focus on conducting inference on a nonlinear function of θ_0. Specifically, we conduct inference on the variance of the quantile treatment effects:

$$\int_0^1 \big(\nabla\theta_0(u)\big)^2\,du - \left(\int_0^1 \nabla\theta_0(u)\,du\right)^2,$$

while imposing that the QTE be increasing in treatment intensity (i.e., $d \mapsto \nabla\theta_0(d)$ is increasing). To map this problem into our framework, we define

$$C_B^m\big([0,1]\big) \equiv \big\{\theta : [0,1] \to \mathbf{R} \text{ s.t. } \|\theta\|_{m,\infty} < \infty\big\}, \qquad \|\theta\|_{m,\infty} \equiv \sup_{0\leq\alpha\leq m}\sup_{d\in[0,1]}\big|\nabla^\alpha\theta(d)\big|,$$

and set $\mathbf{B} = C_B^2([0,1])$ and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$. We impose the restriction that the quantile treatment effect be increasing in the intensity of treatment by letting $\mathbf{G} = C_B^0([0,1])$, $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$ (where we write $\|\cdot\|_\infty$ in place of $\|\cdot\|_{0,\infty}$), and defining

$$\Upsilon_G(\theta) \equiv -\nabla^2\theta. \tag{A.26}$$

As shown in Section A.2.2, $\mathbf{G}$ is a lattice with order unit $\mathbf{1_G} = \mathbf{c}$ for $\mathbf{c}$ the constant function $\mathbf{c}(d) = 1$ for all $d \in [0,1]$. Setting $\mathbf{F} = \mathbf{R}$ with $\|\cdot\|_{\mathbf{F}} = |\cdot|$, we test whether the variance of the quantile treatment effects equals a hypothesized value $\lambda \neq 0$ by setting

$$\Upsilon_F(\theta) = \int_0^1 \big(\nabla\theta(u)\big)^2\,du - \left(\int_0^1 \nabla\theta(u)\,du\right)^2 - \lambda. \tag{A.27}$$

For the parameter space for θ_0, we employ a ball in $\mathbf{B}$ and we thus set Θ to equal

$$\Theta \equiv \big\{\theta \in C_B^2\big([0,1]\big) \text{ s.t. } \|\theta\|_{2,\infty} \leq C_0\big\} \tag{A.28}$$

for some $C_0 < \infty$. For a sequence of approximating functions $\{p_j\}_{j=1}^{j_n}$ defined on $[0,1]$, we then let $p^{j_n}(d) \equiv (p_1(d), \ldots, p_{j_n}(d))'$ and define Θ_n to equal

$$\Theta_n \equiv \big\{p^{j_n\prime}\beta \in C_B^2\big([0,1]\big) : \big\|p^{j_n\prime}\beta\big\|_{2,\infty} \leq C_0\big\}. \tag{A.29}$$

Similarly, for a sequence $\{q_k\}_{k=1}^{k_n}$, we set $q^{k_n}(z) \equiv (q_1(z), \ldots, q_{k_n}(z))'$ and define

$$Q_n(\theta) \equiv \left\|\hat{\Sigma}_n\left\{\frac{1}{n}\sum_{i=1}^n\big(1\{Y_i \leq \theta(D_i)\} - \tau\big)q^{k_n}(Z_i)\right\}\right\|_p$$

for some $2 \leq p \leq \infty$ and weighting matrix $\hat{\Sigma}_n$. The statistics $I_n(R)$ and $I_n(\Theta)$ then equal the minimums of $\sqrt{n}Q_n$ over $\Theta_n \cap R$ and Θ_n, respectively.

In what follows, we will assume for simplicity that θ_0 is identified. As a result, we may think of θ_0 as a function of P through (A.24), though we leave such dependence implicit in the notation. We next impose the following assumptions:

ASSUMPTION A.2.12: *(i)* $\{Y_i, D_i, Z_i\}_{i=1}^n$ *is i.i.d. with* $(Y, D, Z) \in \mathbf{R} \times [0, 1] \times \mathbf{R}$ *distributed according to* $P \in \mathbf{P}$; *(ii) for* Θ *as in* (A.28) *and each* $P \in \mathbf{P}_0$, *there exists a unique* $\theta_0 \in \Theta$ *satisfying* (A.24); *(iii) the distribution of* Y *conditional on* (D, Z) *is absolutely continuous with density* $f_{Y|DZ,P}(\cdot|D, Z)$ *that is bounded and Lipschitz uniformly in* (D, Z) *and* $P \in \mathbf{P}$.

ASSUMPTION A.2.13: *(i) Every* $P \in \mathbf{P}$ *has a density* f_P *w.r.t. Lebesgue measure with compact support* Ω_P *satisfying* $\sup_{P\in\mathbf{P}} \|f_P\|_\infty < \infty$ *and* $\inf_{P\in\mathbf{P}} \inf_{v\in\Omega_P} f_P(v) > 0$; *(ii) for each* $P \in \mathbf{P}$, *there is a continuously differentiable bijection* $T_P : [0, 1]^3 \to \Omega_P$ *with Jacobian* JT_P *and determinant* $|JT_P|$ *satisfying* $\inf_{P\in\mathbf{P}} \inf_{v\in[0,1]^3} |JT_P(v)| > 0$ *and* $\sup_{P\in\mathbf{P}} \sup_{v\in[0,1]^3} \|JT_P(v)\|_o < \infty$.

ASSUMPTION A.2.14: *(i)* $\sup_d \|p^{j_n}(d)\|_2 \lesssim \sqrt{j_n}$; *(ii)* $E_P[p^{j_n}(D)p^{j_n}(D)']$ *has eigenvalues bounded away from zero and infinity uniformly in* $P \in \mathbf{P}$ *and* j_n; *(iii) for each* $P \in \mathbf{P}_0$, *there is a* $\Pi_n\theta_0 \in \Theta_n \cap R$ *satisfying* $\sup_{P\in\mathbf{P}_0} \|E_P[(1\{Y \le \Pi_n\theta_0(D)\} - 1\{Y \le \theta_0(D)\})q^{k_n}(Z)]\|_p = O((n\log(n))^{-1/2})$ *and* $\sup_{P\in\mathbf{P}_0} \|\theta_0 - \Pi_n\theta_0\|_{1,\infty} = o(1)$.

ASSUMPTION A.2.15: *(i)* $\inf_{P\in\mathbf{P}_0} \inf_{\theta\in\Theta:\|\theta-\theta_0\|_{1,\infty}\ge\epsilon} E_P[(P(Y \le \theta(D)|Z) - \tau)^2] > 0$ *for every* $\epsilon > 0$; *(ii) there are* ϵ *and* $\mathrm{s}_n > 0$ *satisfying, for all* $P \in \mathbf{P}_0$ *and* $\|\theta - \Pi_n\theta_0\|_{1,\infty} \le \epsilon$, $\mathrm{s}_n \le \underline{\mathrm{sing}}\{E_P[f_{Y|D,Z}(\theta(D)|D, Z)q^{k_n}(Z)p^{j_n}(D)']\}$ *and* $\mathrm{s}_n = O(1)$.

ASSUMPTION A.2.16: *(i)* $\max_{1\le k\le k_n} \|q_k\|_\infty = O(1)$; *(ii)* $\max_{1\le k\le k_n} \|q_k\|_{1,\infty} = O(k_n)$; *(iii)* $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ *has eigenvalues bounded away from zero and infinity uniformly in* $P \in \mathbf{P}$ *and* k_n; *(iv) for each* $\theta \in \Theta$, *there is a* $\pi_n(\theta) \in \mathbf{R}^{k_n}$ *with* $E_P[(E_P[\rho(X, \theta)|Z] - q^{k_n}(Z)'\pi_n(\theta))^2] = o(1)$ *uniformly in* $P \in \mathbf{P}$ *and* $\theta \in \Theta$; *(v)* $k_n^{1/p}\sqrt{j_n}\log^{3/2}(n)(n^{1/6} \vee k_n)/n^{1/3} = o(1)$ *and* $j_n \log^{3/2}(1 + k_n)k_n^{2/p+1/2}/\mathrm{s}_n\sqrt{n} = o((\log(n))^{-2})$.

ASSUMPTION A.2.17: *(i)* $\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} = o_P((k_n^{1/p}\log(n))^{-1})$ *uniformly in* $P \in \mathbf{P}$; *(ii)* Σ_P *is invertible and* $\|\Sigma_P\|_{o,p}$ *and* $\|\Sigma_P^{-1}\|_{o,p}$ *are bounded in* $P \in \mathbf{P}$ *and* k_n.

Assumptions A.2.12 and A.2.13 impose regularity conditions on the distribution P that enable us to apply an extension of the empirical process coupling results of Koltchinskii (1994). Assumption A.2.14 states the requirements on Θ_n, including demanding an asymptotically negligible bias in Assumption A.2.14(iii). Assumption A.2.15(i) holds pointwise in $P \in \mathbf{P}_0$ due to Θ being compact under $\|\cdot\|_{1,\infty}$, and hence the uniformity in $P \in \mathbf{P}_0$ demanded by Assumption A.2.15(i) corresponds to imposing strong identification. Assumption A.2.15(ii) enables us to obtain a uniform rate of convergence under $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$. As in Section A.2.2, s_n can be shown to be related to the degree of ill-posedness. Assumptions A.2.16(i)–(iv) impose conditions on $\{q_k\}_{k=1}^{k_n}$ including that they be bounded; this requirement can be relaxed at the cost of more stringent rate restrictions to ensure a coupling of the empirical process. Finally, Assumption A.2.16(v) states our rate restrictions, which we note are easier to satisfy for higher values of p.

For any $\theta = p^{j_n\prime}\beta \in \Theta_n \cap R$, in this application the local parameter space equals

$$V_n(\theta, R|\ell) = \biggl\{h = p^{j_n\prime}\beta_h : \Bigl\|\theta + \frac{h}{\sqrt{n}}\Bigr\|_{2,\infty} \le C_0, \sup_{P\in\mathbf{P}} \|h\|_{P,2} \le \ell\sqrt{n},$$

$$\int_0^1 \Bigl(\nabla\theta(u) + \frac{\nabla h(u)}{\sqrt{n}}\Bigr)^2 du - \Bigl(\int_0^1 \Bigl\{\nabla\theta(u) + \frac{\nabla h(u)}{\sqrt{n}}\Bigr\} du\Bigr)^2 = \lambda,$$

$$-\nabla^2\theta(d) - \frac{\nabla^2 h(d)}{\sqrt{n}} \leq 0 \text{ for all } d \in [0,1]\bigg\}, \tag{A.30}$$

where the first two constraints impose that $\theta + h/\sqrt{n} \in \Theta_n$ and $\|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell$, while the final two constraints require that $\theta + h/\sqrt{n} \in R$. Similarly, here

$$V_n(\theta, \Theta|\ell) = \left\{h = p^{j_n\prime}\beta_h : \left\|\theta + \frac{h}{\sqrt{n}}\right\|_{2,\infty} \leq C_0 \text{ and } \sup_{P\in\mathbf{P}} \|h\|_{P,2} \leq \ell\sqrt{n}\right\}.$$

Also recall that $\mathbb{W}_P(\theta) \sim N(0, \mathrm{Var}_P\{\rho(X,\theta)q^{k_n}(Z)\})$, and for any $h = p^{j_n\prime}\beta_h$, define

$$\mathbb{D}_P(\theta)[h] \equiv E_P\big[q^{k_n}(Z)f_{Y|DZ,P}\big(\theta(D)|D,Z\big)p^{j_n}(D)'\beta_h\big]. \tag{A.31}$$

The random variables to which $I_n(R)$ and $I_n(\Theta)$ will be coupled are then given by

$$U_P(R|\ell_n) \equiv \inf_{h\in V_n(\Pi_n\theta_0, R|\ell_n)} \big\|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P(\Pi_n\theta_0)[h]\big\|_{\Sigma_P,2},$$

$$U_P(\Theta|\ell_n) \equiv \inf_{h\in V_n(\Pi_n\theta_0, \Theta|\ell_n)} \big\|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P(\Pi_n\theta_0)[h]\big\|_{\Sigma_P,2}.$$

Our next result obtains distributional approximations by applying Theorem 3.1.

THEOREM A.2.5: *Let Assumptions* A.2.12, A.2.13, A.2.14, A.2.15, A.2.16, *and* A.2.17 *hold*, $a_n = (\log(n))^{-1/2}$, $\ell_n \downarrow 0$ *satisfy* $k_n^{1/p}\sqrt{j_n\ell_n\log(1+k_n)\log(1/\ell_n)} = o((\log(n))^{-1/2})$ *and* $\ell_n^2\sqrt{nj_n\log(n)} = o(1)$. *Then*: (*i*) *Uniformly in* $P \in \mathbf{P}_0$ *it follows that*

$$I_n(R) \leq U_P(R|\ell_n) + o_P(a_n).$$

(*ii*) *If, in addition*, $k_n\log(1+k_n)\sqrt{j_n\log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, *then for any* $\ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $k_n^{1/p}\sqrt{j_n\ell_n^{\mathrm{u}}\log(1+k_n)\log(1/\ell_n^{\mathrm{u}})} = o((\log(n))^{-1/2})$, $(\ell_n^{\mathrm{u}})^2\sqrt{nj_n\log(n)} = o(1)$, *and* $\sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(\ell_n^{\mathrm{u}})$, *it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$I_n(R) - I_n(\Theta) \leq U_P(R|\ell_n) - U_P\big(\Theta|\ell_n^{\mathrm{u}}\big) + o_P(a_n).$$

Theorem A.2.5(i) obtains an upper bound for $I_n(R)$ by relying on Theorem 3.1(i). In order to approximate the re-centered statistic $I_n(R) - I_n(\Theta)$, we cannot rely on an upper bound for $I_n(\Theta)$, as the resulting approximation could fail to control size. Therefore, Theorem A.2.5(ii) instead relies on Theorem 3.1(ii). Applying Theorem 3.1(ii), however, requires an additional rate condition in order to establish the linearization of the moment conditions is asymptotically valid. We also note that the conclusion of Theorem A.2.5(ii) in fact holds with equality if ℓ_n satisfies the same rate restrictions as ℓ_n^{u}.

In order to provide bootstrap estimates for these distributional approximations, we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote minimizers of Q_n over $\Theta_n \cap R$ and Θ_n, respectively. Our bootstrap approximation estimates the law of $\mathbb{W}_P(\theta_0)$ and the derivative $\mathbb{D}_P(\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i\Bigg\{q^{k_n}(Z_i)\big(1\{Y_i \leq \theta(D_i)\} - \tau\big) - \frac{1}{n}\sum_{j=1}^n q^{k_n}(Z_j)\big(1\{Y_j \leq \theta(D_j)\} - \tau\big)\Bigg\},$$

$$\hat{\mathbb{D}}_n(\theta)[h] \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q^{k_n}(Z_i)\left(1\left\{Y_i \leq \theta(D_i) + \frac{h(D_i)}{\sqrt{n}}\right\} - 1\{Y_i \leq \theta(D_i)\}\right),$$

at $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$. An unappealing feature of $\hat{\mathbb{D}}_n(\theta)$ is that it is not linear in h, which complicates computation. Alternatively, a plug-in estimator based on (A.31) could be used, though at the expense of having to estimate the density $f_{Y|DZ,P}$.

With regard to the local parameter space, we note that in this application,

$$G_n(\hat{\theta}_n) \equiv \left\{h = p^{j_n\prime}\beta_h : -\nabla^2\hat{\theta}_n(d) - \frac{\nabla^2 h(d)}{\sqrt{n}} \leq \max\{-\nabla^2\hat{\theta}_n(d) \vee -r_n\} \text{ for all } d \in [0, 1]\right\}.$$

Employing that $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$ and the expression for Υ_F in (A.27), we obtain that

$$\hat{V}_n(\hat{\theta}_n, R|\ell_n) = \left\{h = p^{j_n\prime}\beta_h : h \in G_n(\hat{\theta}_n), \left\|\frac{h}{\sqrt{n}}\right\|_{2,\infty} \leq \ell_n \right.$$
$$\left. \int_0^1 \left(\nabla\hat{\theta}_n(u) + \frac{\nabla h(u)}{\sqrt{n}}\right)^2 du - \left(\int_0^1 \left(\nabla\hat{\theta}_n(u) + \frac{\nabla h(u)}{\sqrt{n}}\right) du\right)^2 = \lambda\right\},$$

where ℓ_n is chosen to satisfy conditions stated below. The bootstrap statistics $\hat{U}_n(R|\ell_n)$ and $\hat{U}_n(\Theta|+\infty)$ for approximating the distributions in Theorem A.2.5 are then

$$\hat{U}_n(R|\ell_n) \equiv \inf_{h \in \hat{V}_n(\hat{\theta}_n, R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\|_{\hat{\Sigma}_n, p},$$
$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{h = p^{j_n\prime}\beta_h} \|\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n(\hat{\theta}_n^{\mathrm{u}})[h]\|_{\hat{\Sigma}_n, p}.$$

The following final assumption will enable us to establish bootstrap validity. In the requirements below, it is helpful to recall θ_0 is implicitly a function of P through (A.24).

ASSUMPTION A.2.18: (*i*) *The functions* $\theta(d) = 1$, $\theta(d) = d^2$ *are in* $\mathbf{B}_n$; (*ii*) $\|\theta_0 - \Pi_n\theta_0\|_{2,\infty} = o(1)$ *uniformly in* $P \in \mathbf{P}_0$ *and* $\sup_{P\in\mathbf{P}_0} \|\theta_0\|_{2,\infty} < C_0$; (*iii*) k_n *satisfies* $k_n^{1/p+12/26} = o(n^{1/26}/\log(n))$; (*iv*) $\sup_d \|\nabla^2 p^{j_n}(d)\|_2 \vee \|\nabla p^{j_n}(d)\|_2 \lesssim j_n^{5/2}$; (*v*) r_n, ℓ_n *satisfy* $k_n^{1/p} \times \sqrt{j_n\ell_n \log(1+k_n)\log(1/\ell_n)} = o((\log(n))^{-1/2})$, $j_n^{5/2}\sqrt{k_n \log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(1 \wedge r_n)$, *and* $\ell_n(\sqrt{j_n n}\ell_n + j_n^{5/2}\sqrt{k_n \log(1+k_n)}/\mathrm{s}_n) = o((\log(n))^{-1/2})$.

Assumption A.2.18(i) requires that the quadratic functions belong to $\mathbf{B}_n$—a condition that holds if quadratic functions belong to the span of $\{p_j\}_{j=1}^{j_n}$. Assumption A.2.18(ii) implies that θ_0 and its approximation $\Pi_n\theta_0$ belong to the interior of Θ. Assumption A.2.18(iii) enables us to verify the bootstrap coupling requirement of Assumption 3.11. While condition A.2.18(iii) suffices for verifying Assumption 3.11 in both the endogenous ($Z \neq D$) and exogenous ($Z = D$) settings, we note that, in both cases, better rate conditions can be obtained.[1] Finally, Assumption A.2.18(iv) ensures $\mathcal{S}_n(\mathbf{B}, \mathbf{E}) \asymp j_n^{5/2}$, while Assumption A.2.18(v) imposes the requirements on ℓ_n and r_n.

The next theorem establishes the validity of the bootstrap procedure.

[1]For instance, under endogeneity, a better rate could be obtained by conducting a basis expansion using the tensor product of a Haar Basis for (Y, D) and the functions $\{q_k\}_{k=1}^{k_n}$.

THEOREM A.2.6: *Let Assumptions A.2.12, A.2.13, A.2.14, A.2.15, A.2.16, A.2.17, and A.2.18 hold and* $a_n = (\log(n))^{-1/2}$. *Then, there is a sequence* $\tilde{\ell}_n \asymp \ell_n$ *satisfying*

$$\hat{U}_n(R|\ell_n) \geq U_P^{\star}(R|\tilde{\ell}_n) + o_P(a_n)$$

uniformly in $P \in \mathbf{P}_0$. (*ii*) *If, in addition,* $k_n \log(1 + k_n)\sqrt{j_n \log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, *then for any* $\tilde{\ell}_n^{\mathrm{u}}$ *satisfying the conditions of Theorem A.2.5*(*ii*), *we have, uniformly in* $P \in \mathbf{P}_0$,

$$\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) \geq U_P^{\star}(R|\tilde{\ell}_n) - U_P^{\star}(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n).$$

Theorems A.2.5(i) and A.2.6(i) imply that as critical value for $I_n(R)$, we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n)) \equiv \inf\{c : P(\hat{U}_n(R|\ell_n) \leq c|\{V_i\}_{i=1}^n) \geq 1 - \alpha\}.$$

If, in addition, $k_n \log(1 + k_n)\sqrt{j_n \log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, then Theorems A.2.5(ii) and A.2.6(ii) imply a valid test can be obtained by rejecting whenever $I_n(R) - I_n(\Theta)$ exceeds

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty)) \equiv \inf\{c : P(\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) \leq c|\{V_i\}_{i=1}^n) \geq 1 - \alpha\}.$$

Our critical values depend on the choices of r_n and ℓ_n. The slackness parameter r_n again measures sampling uncertainty in whether constraints "bind." Following the discussion in Section 2.1, for $\hat{\theta}_n^{\mathrm{u}\star}$ a "bootstrap" analogue to $\hat{\theta}_n^{\mathrm{u}}$, we may thus set

$$P\Big(\max_{d\in[0,1]} \nabla^2\hat{\theta}_n^{\mathrm{u}}(d) - \nabla^2\hat{\theta}_n^{\mathrm{u}\star}(d) \leq r_n|\{V_i\}_{i=1}^n\Big) = 1 - \gamma_n, \tag{A.32}$$

with $\gamma_n \to 0$. With regard to ℓ_n, we note that its main role in this application is to ensure that $\hat{V}_n(\hat{\theta}_n, R|\ell_n)$ is well approximated by the true local parameter space despite the nonlinearity of Υ_F. To this end, the requirements on ℓ_n imposed in Assumption A.2.18 ensure $\sqrt{n}\ell_n\|\hat{\theta}_n - \Pi_n\theta_0\|_{\mathbf{B}} = o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$. Since $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$ in this application, we may select ℓ_n in a data driven way by setting it to satisfy

$$P\Big(\max_{d\in[0,1]}|\nabla^2\hat{\theta}_n^{\mathrm{u}}(d) - \nabla^2\hat{\theta}_n^{\mathrm{u}\star}(d)| \leq \frac{1}{\sqrt{n}\ell_n}|\{V_i\}_{i=1}^n\Big) = 1 - \gamma_n \tag{A.33}$$

for some $\gamma_n \to 0$. While we set γ_n in (A.32) and (A.33) to be the same, it is worth noting they could be different. In fact, r_n and ℓ_n do not "interact" in the requirements of Assumption A.2.18(v) and, in this sense, can be set independently. We also note that in settings in which the rate of convergence is sufficiently fast, (A.33) should deliver a "large" ℓ_n in the sense that $\hat{U}_n(R|\ell_n)$ and $\hat{U}_n(R| + \infty)$ are asymptotically equivalent. Moreover, in applications in which we expect the rate of convergence of $\hat{\theta}_n$ to be sufficiently fast, we may directly set $\ell_n = +\infty$.

REMARK A.2.3: To illustrate the role of ℓ_n, it is helpful to conduct a pointwise (in P) analysis, set $p = 2$, and connect our assumptions to the literature on estimation of conditional moment restriction models (Chen and Pouzo (2012)). We follow the literature in imposing a local curvature assumption, which in our application corresponds to

$$\begin{aligned}&\|E_P[(P(Y \leq h(D)|Z) - \tau)q^{k_n}(Z)]\|_2 \\ &\quad \asymp \|E_P[f_{Y|DZ,P}(\bar{\theta}(D)|D, Z)(\theta_0(D) - h(D))q^{k_n}(Z)]\|_2\end{aligned} \tag{A.34}$$

for all $h \in \Theta_n$ and $\bar{\theta} \in \Theta$ that are in a neighborhood of θ_0. We further suppose the linear operator $h \mapsto E_P[f_{Y|DZ,P}(\theta_0(D)|D, Z)h(D)|Z]$ is compact, in which case there exist orthonormal bases $\{\psi_j\}$ and $\{\phi_k\}$ and a sequence $\lambda_j \downarrow 0$ satisfying

$$E_P\big[f_{Y|DZ,P}\big(\theta_0(D)|D, Z\big)\phi_j(D)|Z\big] = \lambda_j \psi_j(Z). \tag{A.35}$$

Setting $k_n \geq j_n$ with $k_n - j_n = O(1)$, $p^{j_n} = (\phi_1, \ldots, \phi_{j_n})'$, $q^{k_n} = (\psi_1, \ldots, \psi_{k_n})'$, and $\Pi_n^{\mathrm{u}}\theta_0 = \sum_{j=1}^{j_n} \phi_j \beta_j$, we also suppose θ_0 admits an expansion

$$\theta_0 = \sum_{j=1}^{\infty} \beta_j \phi_j \quad \text{with } |\beta_j| = O\big(j^{-\gamma_\beta}\big). \tag{A.36}$$

Provided that the approximation error of $\Pi_n\theta_0$ (as in Assumption A.2.14(iii)) and $\Pi_n^{\mathrm{u}}\theta_0$ are of the same order, it then follows from (A.34) and (A.35) that

$$\big\| E_P\big[\big(1\big\{Y \leq \Pi_n\theta_0(D)\big\} - 1\big\{Y \leq \theta_0(D)\big\}\big)q^{k_n}(Z)\big]\big\|_2 \lesssim \frac{\lambda_{j_n}}{j_n^{\gamma_\beta}} \tag{A.37}$$

and $\mathrm{s}_n \asymp \lambda_{j_n}$—that is, s_n is proportional to the reciprocal of the sieve measure of ill-posedness (Chen and Pouzo (2012)). As a result, if $\lambda_j \asymp j^{-\gamma_\lambda}$ and $\gamma_\beta > \max\{5/2, 3 - \gamma_\lambda\}$, then Theorem A.2.5 may be applied to couple $I_n(R)$ by setting $j_n \asymp n^\kappa$ with $(2(\gamma_\lambda + \gamma_\beta))^{-1} < \kappa < \min\{(5 + 2\gamma_\lambda)^{-1}, 1/6\}$, while coupling $I_n(R) - I_n(\Theta)$ additionally requires $\gamma_\beta > 3/2 + \gamma_\lambda$ and $\kappa < (3 + 4\gamma_\lambda)^{-1}$. In contrast, in the severely ill-posed case in which $\lambda_j \asymp \exp\{-\gamma_\lambda j\}$, the conditions of Theorem A.2.5 for coupling $I_n(R) - I_n(\Theta)$ are not satisfied. However, the conditions for coupling $I_n(R)$ can still be met provided $\gamma_\beta > 4$ by setting $j_n = (\log(n) - \kappa(\log(\log(n))))/2\gamma_\lambda$ with $7 < \kappa < 2\gamma_\beta - 1$. Thus, while in the ill-posed case the rate of convergence is too slow to apply Theorem A.2.5(ii), Theorem A.2.5(i) is still able to deliver a coupling upper bound for suitable ℓ_n.

REFERENCES

ALIPRANTIS, CHARALAMBOS D., AND KIM C. BORDER (2006): *Infinite Dimensional Analysis—a Hitchhiker's Guide*. Berlin: Springer-Verlag. [1]

BELLONI, ALEXANDRE, VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND KENGO KATO (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345–366. [8]

BLUNDELL, RICHARD, XIAOHONG CHEN, AND DENNIS KRISTENSEN (2007): "Semi-Nonparametric iv Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669. [11]

BLUNDELL, RICHARD, JOEL L. HOROWITZ, AND MATTHIAS PAREY (2012): "Measuring the Price Responsiveness of Gasoline Demand: Economic Shapre Restrictions and Nonparametric Demand Estimation," *Quantitative Economics*, 3, 29–51. [6]

——— (2017): "Nonparametric Estimation of a Nonseparable Demand Function Under the Slutsky Inequality Restriction," *Review of Economics and Statistics*, 99, 291–304. [6]

CHEN, XIAOHONG (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, Vol. 6B, ed. by James J. Heckman and Edward E. Leamer. North Holland, Elsevier. [8]

CHEN, XIAOHONG, AND TIMOTHY M. CHRISTENSEN (2018): "Optimal sup-Norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric iv Regression," *Quantitative Economics*, 9, 39–84. [8]

CHEN, XIAOHONG, AND DEMIAN POUZO (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80, 277–321. [16,17]

CHERNOZHUKOV, VICTOR, AND CHRISTIAN HANSEN (2005): "An iv Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. [12]

CHERNOZHUKOV, VICTOR, WHITNEY K. NEWEY, AND ANDRES SANTOS (2022): "Constrained Conditional Moment Restriction Models," arXiv preprint, arXiv:1509.06311. [1]

HANSEN, LARS P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 891–916. [2]

KOLTCHINSKII, VLADIMIR I. (1994): "Komlos–Major–Tusnady Approximation for the General Empirical Process and Haar Expansions of Classes of Functions," *Journal of Theoretical Probability*, 7, 73–118. [13]

MATZKIN, ROSA L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Vol. IV, ed. by Robert Engle and Dan McFadden. Elsevier. [6]

NEWEY, WHITNEY K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [8]

YURINSKII, VADIM V. (1977): "On the Error of the Gaussian Approximation for Convolutions," *Theory of Probability and Its Applications*, 2, 236–247. [4]

ZHAI, ALEX (2018): "A High-Dimensional clt in w_2 Distance With Near Optimal Convergence Rate," *Probability Theory and Related Fields*, 170, 821–845. [8]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 24 September, 2015; final version accepted 2 August, 2022; available online 4 November, 2022.