Supplementary Material

Supplement to "Inference in nonparametric/semiparametric moment equality models with shape restrictions"

(*Quantitative Economics*, Vol. 11, No. 2, May 2020, 609–636)

Yu Zhu
Bank of Canada

This supplemental material has the following content: (1) proofs of all the theoretic results in the main paper; (2) a set of low-level conditions that imply Definition 5.1(iii) in the main paper; (3) an example where point-identification is achieved with a shape restriction, while one cannot rule out any point in the parameter space without the shape restriction; (4) examples to illustrate that either my method nor CNS dominates the other.

I start with introducing some additional notation used in this supplemental material. Let $\{\delta_n\}_{n=1}^{\infty}$ be a sequence of nonnegative numbers and $X_{n,F}$ be a sequence of random variables indexed by n and F. Say $X_{n,F} = O_{\mathbb{P}_F}(\delta_n)$ uniformly in $F \in \mathcal{F}$ if and only if (iff) for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F(|X_{n,F}| > M\delta_n) < \epsilon$ and $X_{n,F} = o_{\mathbb{P}_F}(\delta_n)$ uniformly in $F \in \mathcal{F}$ iff for all $M > 0$, $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F(|X_{n,F}| > M\delta_n) = 0$. Similarly, $X_{n,F,R} = O_{\mathbb{P}_F}(\delta_n)$ uniformly in $(F,R) \in \mathcal{J}$ iff for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F(|X_{n,F,R}| > M\delta_n) < \epsilon$ and $X_{n,F,R} = o_{\mathbb{P}_F}(\delta_n)$ uniformly in $(F,R) \in \mathcal{J}$ iff for all $M > 0$, $\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F(|X_{n,F,R}| > M\delta_n) = 0$. Because $(F,R) \in \mathcal{J}$ implies $F \in \mathcal{F}$, $X_{n,F} = O_{\mathbb{P}_F}(\delta_n)$ ($X_{n,F} = o_{\mathbb{P}_F}(\delta_n)$) uniformly in $(F,R) \in \mathcal{J}$ if $X_{n,F} = O_{\mathbb{P}_F}(\delta_n)$ ($X_{n,F} = o_{\mathbb{P}_F}(\delta_n)$) uniformly in $F \in \mathcal{F}$. Lastly, X_n is $O_{\mathbb{P}^*}(\delta_n)$, $\mathbb{P}_F$ almost surely iff given any $\epsilon > 0$, for almost every sample realization under $\mathbb{P}_F$, there exists an $M > 0$ such that $\lim_{n\to\infty} \mathbb{P}_{F_n}(X_n > M\delta_n) < \epsilon$, where $\mathbb{P}_{F_n}$ is induced by the empirical distribution. Similarly, X_n is $o_{\mathbb{P}^*}(\delta_n)$, $\mathbb{P}_F$ almost surely iff for almost every sample realization under $\mathbb{P}_F$, $\lim_{n\to\infty} \mathbb{P}_{F_n}(X_n > M\delta_n) = 0$ for all $M > 0$.

For any function $g : \ell^\infty(\Theta \times \mathbf{T}) \to \mathbb{R}$, let $q(1-\alpha, g, F)$ be the $1-\alpha$th quantile of $g(\mathbb{G}_F)$ and $q_n^*(1-\alpha, g)$ be the $1-\alpha$th quantile of $g(\mathbb{G}_n^*)$. And $\mathbf{1}(\cdot)$ is the indicator function which takes value 1 if the event in the bracket is true and value 0 if otherwise.

Other notation can be found in Table S.1, and the text and Appendix A of the main paper. Assumptions B.1–B.7 can be found in Appendix B of the main paper.

S.1. Identification power of shape restrictions

It is commonly known that shape restrictions can play an important role for identification in structural models, for example, in BLP estimation and auctions; see Gandhi, Lu, and Shi (2013), Zhu and Grundl (2014), Komarova (2013), and Fan, He, and Li (2015). But

Yu Zhu: zhuyuzlf57@gmail.com

Available at http://qeconomics.org. https://doi.org/10.3982/QE1023

TABLE S.1. Some additional notation.

$\|\cdot\|_E$	Euclidean norm
$\|\cdot\|_{2,F}$	L^2 norm under distribution F, that is, $\|f\|_{2,F}^2 = \int \|f(w)\|_E^2\, dF(w)$
$\vec{d}_F(A, B)$	$\sup_{a\in A} \inf_{b\in B} d_F(a, b)$, directed Hausdorff metric between $A, B \subset \Theta$
$\vec{d}_{w,F}(A, B)$	$\sup_{a\in A} \inf_{b\in B} d_{w,F}(a, b)$ for $A, B \subset \Theta$
$R^{\epsilon}_{n,F}(\theta)$	$\Theta_n \cap R \cap B^{\epsilon}_F(\theta)$
$\mathbb{E}^*$	Expectation under the empirical distribution
$\Pi_{n,R}\theta$	Projection of θ on to $\Theta_n \cap R$, that is, $\arg\min_{\tilde{\theta}\in\Theta_n\cap R} d_s(\tilde{\theta}, \theta)$
$\uparrow$	Converge from below
$\to^d$	Converge in distribution to

even in the simple non-parametric IV regression model, shape restrictions can greatly help identification. In extreme cases, without shape restrictions, the parameter of interest is not identified at all, that is, one cannot rule out any point from the parameter space. But once a shape restriction is imposed, the parameter of interest is point-identified. This section provides such an example.

Consider the setup in Example 2.1 in the main paper where the joint distribution of (X, Y, Z) is such that

$$Z \sim \text{Uniform}[0, 1], \qquad X|Z \sim \text{Uniform}[Z, Z + 1].$$

And ϵ can be arbitrarily correlated with X but satisfies $\mathbb{E}_F(\epsilon|Z) = 0$. For example, $\epsilon = Z(X - Z + \epsilon_1 - 1/2)$, where ϵ_1 is a random variable with mean 0 and independent of both X and Z. From this point on, I restrict θ_F to be a bounded continuous function and the parameter of interest is $\theta_F(x_0)$ for some $x_0 \in [0, 2]$.

Suppose that one uses the conditional moment restriction

$$\mathbb{E}_F\big[Y - \theta(X)|Z\big] = 0. \tag{S.1}$$

Then the identified set of $\theta_F(x_0)$ is $(-\infty, \infty)$. To see this, notice that $\theta = \theta_F + \tilde{\theta}$ satisfies equation (S.1) iff $\mathbb{E}_F[\tilde{\theta}(X)|Z] = 0$, which holds iff $\tilde{\theta}$ is a periodic function with period 1 and $\int_0^1 \tilde{\theta}(x)\, dx = 0$. For any real number a, there exists such a $\tilde{\theta}$ with $\tilde{\theta}(x_0) = a$. Hence, $\theta_F(x_0)$ can take any value in $(-\infty, \infty)$.

Now suppose θ_F is weakly increasing and is constant on $[b, c] \subseteq [0, 2]$ with $c - b \geq 1$. After imposing the weakly increasing restriction, $\theta_F(x_0)$ is point-identified. To see this, first notice $\tilde{\theta}$ has to be a constant function. Otherwise, $\tilde{\theta}$ must decrease in some region of $[b, c]$ because it has period 1, which means that $\theta_F + \tilde{\theta}$ is not weakly increasing. Then $\int_0^1 \tilde{\theta}(x)\, dx = 0$ suggests that $\tilde{\theta} = 0$ and θ_F is point-identified.

S.2. CHOICES OF Θ AND LOW-LEVEL SUFFICIENT CONDITIONS FOR DEFINITION 5.1(III)

Define $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_{D_W})$ to be a D_W-dimensional vector of nonnegative integers and $w = (w_1, w_2, \ldots, w_{D_W}) \in \mathcal{W}$ to be a D_W-dimensional vector of real numbers. Let $|\Lambda| = \sum_{i=1}^{D_W} \Lambda_i$ and $d^{\Lambda} f = \partial^{|\Lambda|} f(w)/\partial^{\Lambda_1} w_1 \cdots \partial^{\Lambda_{D_W}} w_{D_W}$ where $f : \mathcal{W} \to \mathbb{R}$ is a function.

Introduce the norms

$$\|f\|_c = \max_{|\Lambda|\le m} \sup_{w\in\mathcal{W}} \left|d^\Lambda f(w)\right|\left(1+ww'\right)^{\delta/2},$$

$$\|f\|_{\text{sob}}^2 = \int \max_{|\Lambda|\le m+m_0} \left|d^\Lambda f(w)\right|^2\left(1+ww'\right)^{\delta_0} dw,$$

where m and m_0 are two integers. Let $\mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W}) = \{f : \mathcal{W} \to \mathbb{R} \text{ s.t. } \|f\|_{\text{sob}} \le \mathbf{B}\}$ where $0 < \mathbf{B} < \infty$ is a known constant. Define $\Theta = \Theta^P \times \mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W})^{D_N}$ where Θ^P is a compact subset of $\mathbb{R}^{D_P}$ and D_N is a positive integer. In addition, let $\theta = (\theta^P, \theta^N)$ where $\theta^P \in \Theta^P$ and $\theta^N \in \mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W})^{D_N}$. Define

$$\begin{aligned}\|\theta\|_s = \|\theta\|_\infty &= \max\left\{\left\|\theta^P\right\|_{\infty,E}, \sup_{w\in\mathcal{W}}\left\|\theta^N(w)\right\|_{\infty,E}\right\}\\ &= \max\left\{\left\|\theta^P\right\|_{\infty,E}, \sup_{w\in\mathcal{W}}\left|\theta_1^N(w)\right|, \sup_{w\in\mathcal{W}}\left|\theta_2^N(w)\right|, \dots, \sup_{w\in\mathcal{W}}\left|\theta_{D_N}^N(w)\right|\right\},\end{aligned}$$

where $\|\cdot\|_{\infty,E}$ is the sup norm in the Euclidean space and $\theta_j^N(w)$ is the jth element of $\theta^N(w)$. The following is Assumption 2.1 in Santos (2012).

Assumption S.1. (i) *For $\mathcal{W}$ bounded, $\delta_0 = \delta = 0$ and $\min\{m_0, m\} > D_W/2$, while for $\mathcal{W}$ unbounded, $m_0 > D_W/2$ and $D_W(m+\delta)/(m\delta) < 2$ and $\delta_0 > \delta > D_W/2$.* (ii) *$\mathcal{W}$ satisfies a uniform cone condition.*

Santos (2012) showed that under Assumption S.1, $\mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W})$ is compact under $\|\cdot\|_c$. Since $\|\cdot\|_c$ is stronger than the sup norm, $\mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W})$ is compact under the sup norm. Therefore, Θ is compact under $\|\cdot\|_s$ as it is a product of compact spaces. Moreover, suppose that the shape restriction is

$$R = \left\{\theta : \partial\theta_1^N(w)/\partial w_1 \ge 0, \forall w \in \mathcal{W}\right\}.$$

Because partial differentiation operators are continuous with respect to $\|\cdot\|_c$, R is closed under $\|\cdot\|_c$. This implies that $\Theta \cap R$ is compact under $\|\cdot\|_c$, and therefore compact under $\|\cdot\|_s$. If R involves higher-order derivatives, one needs to adjust Θ accordingly to satisfy Assumption B.1(i). Suppose $\mathbf{t} \in \mathbf{T}$ is a $D_{\mathbf{t}}$-dimensional real vector. Define $\mathbf{N}(\epsilon, \Theta \times \mathbf{T}, \|\cdot\|_s + \|\cdot\|_E)$ to be the smallest ϵ-covering number of $\Theta \times \mathbf{T}$ under $\|\cdot\|_s + \|\cdot\|_E$.

Lemma S.1. *Under Assumption S.1, if $\mathbf{T}$ is a bounded subset of $\mathbb{R}^{D_{\mathbf{t}}}$, then*

$$\int_0^\infty \sqrt{\ln \mathbf{N}\left(\epsilon, \Theta \times \mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\right)}\, d\epsilon < \infty.$$

Proof. Lemma A.3 in Santos (2012) shows that

$$\ln \mathbf{N}\left(\epsilon, \mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W}), \|\cdot\|_\infty\right) \le K_1\left(\frac{1}{\epsilon}\right)^{D_W(m+\delta)/(\delta m)} \quad \text{if } \mathcal{W} \text{ unbounded},$$

$$\ln \mathbf{N}\left(\epsilon, \mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W}), \|\cdot\|_\infty\right) \le K_2\left(\frac{1}{\epsilon}\right)^{D_W/m} \quad \text{if } \mathcal{W} \text{ bounded},$$

where $\mathbf{N}(\epsilon, \mathcal{E}_B^{\text{sob}}(\mathcal{W}), \|\cdot\|_\infty)$ is the smallest ϵ-covering number of $\mathcal{E}_B^{\text{sob}}(\mathcal{W})$ under $\|\cdot\|_\infty$. Notice that

$$\mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big) < \mathbf{N}\left(\frac{\epsilon}{2}, \Theta, \|\cdot\|_s\right)\mathbf{N}\left(\frac{\epsilon}{2}, \mathbf{T}, \|\cdot\|_E\right)$$
$$\leq \mathbf{N}\left(\frac{\epsilon}{2}, \Theta^P, \|\cdot\|_{\infty,E}\right)\mathbf{N}\left(\frac{\epsilon}{2}, \mathbf{T}, \|\cdot\|_E\right)\mathbf{N}\left(\frac{\epsilon}{2}, \mathcal{E}_{\mathbf{B}}^{\text{sob}}(\mathcal{W}), \|\cdot\|_\infty\right)^{D_N}.$$

Therefore, there exists a constant K such that

$$\ln\mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big) \leq K\left(\frac{1}{\epsilon}\right)^{D_W(m+\delta)/(\delta m)} \quad \text{if } \mathcal{W} \text{ unbounded,}$$
$$\ln\mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big) \leq K\left(\frac{1}{\epsilon}\right)^{D_W/m} \quad \text{if } \mathcal{W} \text{ bounded.}$$

Let $M = \max_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}(\|\theta\|_s + \|\mathbf{t}\|_E)$. Then under Assumption S.1,

$$\int_0^\infty \sqrt{\ln\mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big)}\, d\epsilon \leq K\int_0^M \left(\frac{1}{\epsilon}\right)^\eta d\epsilon,$$

where $\eta = D_W(m+\delta)/(2\delta m)$ if $\mathcal{W}$ is unbounded and $\eta = D_W/2m$ if $\mathcal{W}$ is bounded. Because $\eta < 1$ by Assumption S.1, the right-hand side equals $KM^{1-\eta}/(1-\eta) < \infty$. □

ASSUMPTION S.2. *$\rho_\mathbf{t}$ and $\mathcal{F}$ satisfy* (i) *$\rho_\mathbf{t}(\cdot,\theta)$ is measurable for all $(\theta,\mathbf{t}) \in \Theta\times\mathbf{T}$.* (ii) *$\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\rho_\mathbf{t}(\cdot,\theta)| \leq \mathbf{F}(\cdot)$ for some function $\mathbf{F}$.* (iii) *There exists $\mathbf{L}(\cdot)$ such that $\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{L}(W_i)^2 < \infty$ and*

$$\big|\rho_{\mathbf{t}_1}(\cdot,\theta_1) - \rho_{\mathbf{t}_2}(\cdot,\theta_2)\big| \leq \mathbf{L}(\cdot)\big[\|\theta_1-\theta_2\|_s + \|\mathbf{t}_1-\mathbf{t}_2\|_E\big].$$

(iv) *$\limsup_{M\uparrow\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{F}(W_i)^2\mathbf{1}\{\mathbf{F}(W_i) > M\} = 0$.*

Assumption S.2(iii) is a uniform version of the standard smoothness assumption on the moment functions. Assumption S.2(iv) is a uniform integrability condition, which holds if $\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{F}(W_i)^{2+\delta} < \infty$ for some $\delta > 0$.

LEMMA S.2. *Under Assumptions S.1 and S.2, if $\mathbf{T}$ is a bounded subset of $\mathbb{R}^{D_\mathbf{t}}$, ϱ is Donsker and pre-Gaussian uniformly in $F\in\mathcal{F}$.*

PROOF. By Assumption S.2(iii),

$$\big\|\rho_{\mathbf{t}_1}(\cdot,\theta_1) - \rho_{\mathbf{t}_2}(\cdot,\theta_2)\big\|_{2,F}^2 \leq \mathbb{E}_F\mathbf{L}(W_i)^2\big[\|\theta_1-\theta_2\|_s + \|\mathbf{t}_1-\mathbf{t}_2\|_E\big]^2.$$

Let $\mathbf{N}_{[]}$ be the bracketing number. By Theorem 2.7.11 in Van Der Vaart and Wellner (1996),

$$\mathbf{N}_{[]}\big(2\epsilon(\|\mathbf{L}\|_{2,F}+1), \varrho, \|\cdot\|_{2,F}\big) \leq \mathbf{N}_{[]}\big(2\epsilon\|\mathbf{L}\|_{2,F}, \varrho, \|\cdot\|_{2,F}\big) \leq \mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big).$$

Without loss of generality, assume that $\mathbf{F}(\cdot) \geq 1$. This is because one can always redefine the envelope function to be $\max\{\mathbf{F}(\cdot), 1\}$. Then

$$
\begin{aligned}
&\int_0^\infty \sup_{F\in\mathcal{F}} \sqrt{\ln \mathbf{N}_{[]}\big(\epsilon\|\mathbf{F}\|_{2,F}, \varrho, \|\cdot\|_{2,F}\big)}\, d\epsilon \\
&\quad \leq \int_0^\infty \sup_{F\in\mathcal{F}} \sqrt{\ln \mathbf{N}\bigg(\frac{\epsilon\|\mathbf{F}\|_{2,F}}{2\big(\|\mathbf{L}\|_{2,F}+1\big)}, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\bigg)}\, d\epsilon \\
&\quad \leq \sup_{F\in\mathcal{F}} \int_0^\infty \frac{2\big(\|\mathbf{L}\|_{2,F}+1\big)}{\|\mathbf{F}\|_{2,F}} \sqrt{\ln \mathbf{N}\big(\epsilon, \Theta\times\mathbf{T}, \|\cdot\|_s + \|\cdot\|_E\big)}\, d\epsilon,
\end{aligned}
$$

which is bounded by Lemma S.1. Also by Assumption S.2, $\lim_{M\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{F}^2 \times \mathbf{1}(\mathbf{F} > M) = 0$. Then Theorem 2.8.4 in Van Der Vaart and Wellner (1996) implies that ϱ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{F}$. □

S.3. Asymptotic approximation of the test statistic

Theorem S.1. *If Assumptions* B.1 *to* B.4 *hold, then*:

(i) *Uniformly in* $(F, R) \in \mathcal{J}$, $T_n(R) \leq \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_{\mathbb{P}_F}(1)$, *where the inequality holds as an equality if* $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = o_{\mathbb{P}_F}(\xi_n)$ *uniformly in* $(F, R) \in \mathcal{J}$.

(ii) *If* $F \in \mathcal{F}$ *and* $\Theta_F \cap R = \emptyset$, $T_n/n \to \min_{\theta\in\Theta\cap R} Q_F(\theta)$, $\mathbb{P}_F$ *almost surely.*

Proof. For any $(F, R) \in \mathcal{J}$,

$$
\begin{aligned}
T_n(R) &= \inf_{\theta\in\Theta_n\cap R} \int_{\mathbf{T}} \bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^n \big[\rho_{\mathbf{t}}(W_i,\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\big] + \sqrt{n}\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta) \bigg|^2 d\mu(\mathbf{t}) \\
&\leq \inf_{\tilde{\theta}\in\Theta_{n,F}(R)} \inf_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \int_{\mathbf{T}} \big|\mathbb{G}_{n,F}(\theta,t) + \sqrt{n}\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\big|^2 d\mu(\mathbf{t}) \\
&= \inf_{\tilde{\theta}\in\Theta_{n,F}(R)} \inf_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \int_{\mathbf{T}} \bigg|\mathbb{G}_{n,F}(\tilde{\theta},\mathbf{t}) + \sqrt{n}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}] + \Delta_n(\mathbf{t})\bigg|^2 d\mu(\mathbf{t}) \\
&\equiv \overline{T}_n(R),
\end{aligned} \tag{S.2}
$$

where $\Delta_n(\mathbf{t}) = \Delta_{n,1}(\mathbf{t}) + \Delta_{n,2}(\mathbf{t}) + \Delta_{n,3}(\mathbf{t})$ with

$$
\begin{aligned}
\Delta_{n,1}(\mathbf{t}) &= \mathbb{G}_{n,F}(\theta,\mathbf{t}) - \mathbb{G}_{n,F}(\tilde{\theta},\mathbf{t}), \\
\Delta_{n,2}(\mathbf{t}) &= \sqrt{n}\bigg(\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta}) - \frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\bigg), \\
\Delta_{n,3}(\mathbf{t}) &= \sqrt{n}\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta}).
\end{aligned}
$$

I suppress the dependence of the Δs on F, R, θ, and $\tilde{\theta}$ to simplify notation. The inequality in equation (S.2) holds because I shrink the region over which the inf is taken. By the triangular inequality,

$$\sqrt{\overline{T}_n(R)} \le \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} + \sum_{j=1}^{3} \sup_{\tilde{\theta}\in\Theta_{n,F}(R)} \sup_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \left\|\Delta_{n,j}(\cdot)\right\|_{2,\mu},$$

where $\|\Delta_{n,j}(\cdot)\|^2_{2,\mu} = \int_{\mathbf{T}} \Delta_{n,j}(\mathbf{t})^2\, d\mu(\mathbf{t})$. Definition 5.1(iii), Assumptions B.4(iii), and Theorem 2.8.2 in Van Der Vaart and Wellner (1996) imply that

$$\sup_{\tilde{\theta}\in\Theta_{n,F}(R)} \sup_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \|\Delta_{n,1}\|_{2,\mu} \le \sup_{d_F(\theta_1,\theta_2)\le\xi_n} \sup_{\mathbf{t}\in\mathbf{T}} \left|\mathbb{G}_{n,F}(\theta_1,\mathbf{t}) - \mathbb{G}_{n,F}(\theta_2,\mathbf{t})\right| = o_{\mathbb{P}_F}(1),$$

uniformly in $F\in\mathcal{F}$. In addition, under Assumptions B.4(i) and (ii), uniformly in $(F,R)\in\mathcal{J}$

$$\sup_{\tilde{\theta}\in\Theta_{n,F}(R)} \sup_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \|\Delta_{n,2}\|_{2,\mu} \le \sup_{\tilde{\theta}\in\Theta_{n,F}(R)} \sup_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \sup_{\mathbf{t}\in\mathbf{T}} \left|\Delta_{n,2}(\mathbf{t})\right| = o(1).$$

Lastly, because $\tilde{\theta}\in\Theta_{n,F}(R)$, Assumption B.1(ii) implies that for any $\dot{\theta}\in\Theta_F\cap R$,

$$\|\Delta_{n,3}\|^2_{2,\mu} = n \min_{\theta\in\Theta_n\cap R} Q_F(\theta) \le nQ_F(\Pi_{n,R}\dot{\theta}) = nd_{w,F}(\Pi_{n,R}\dot{\theta},\dot{\theta})^2 \le nC_2 d_s(\Pi_{n,R}\dot{\theta},\dot{\theta})^2.$$

Then Assumption B.2(iii) and the definition of $\Pi_{n,R}\dot{\theta}$ imply that uniformly in $(F,R)\in\mathcal{J}$,

$$\sup_{\tilde{\theta}\in\Theta_{n,F}(R)} \sup_{\theta\in R^{\xi_n}_{n,F}(\tilde{\theta})} \|\Delta_{n,3}\|_{2,\mu} = o(1).$$

Therefore, $\sqrt{\overline{T}_n(R)} \le \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} + o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$. Similarly, one can show that $\sqrt{\overline{T}_n(R)} \ge \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} - o_{\mathbb{P}_F}(1)$. Therefore, $\sqrt{\overline{T}_n(R)} = \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} + o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$. Because $\tilde{\theta}\in R^{\xi_n}_{n,F}(\tilde{\theta})$, $\sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} = O_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$. Apply Lemma S.10 with $\mathcal{A}=\mathcal{J}$ to get $\overline{T}_n(R) = \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$, which implies that $T_n(R) \le \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$. If, in addition, $\vec{d}_F(\hat{\Theta}_n(R),\Theta_{n,F}(R)) = o_{\mathbb{P}_F}(\xi_n)$ uniformly in $(F,R)\in\mathcal{J}$, $T_n(R) = \overline{T}_n(R)$ with probability approaching 1 uniformly in $(F,R)\in\mathcal{J}$. Then $T_n(R) = \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$.

For the second claim, notice that by definition, ϱ is Glivenko–Cantelli under any $F\in\mathcal{F}$. Therefore, $\mathbb{P}_F$ almost surely,

$$\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} \left|\bar{\rho}_{\mathbf{t}}(\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\right| = o(1),$$

where $\bar{\rho}_{\mathbf{t}}(\theta) = n^{-1}\sum_{i=1}^n \rho_{\mathbf{t}}(W_i,\theta)$. Because $\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)$ is continuous in θ and $\Theta\cap R$ is compact under d_s, $\mathbb{P}_F$ almost surely,

$$\sqrt{\frac{T_n(R)}{n}} \ge \inf_{\theta\in\Theta_n\cap R} \sqrt{\int_{\mathbf{T}} \left|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\right|^2 d\mu(\mathbf{t})} - o(1) \ge \sqrt{\min_{\theta\in\Theta\cap R} \int_{\mathbf{T}} \left|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\right|^2 d\mu(\mathbf{t})} - o(1).$$

Similarly, by the compactness of $\Theta \cap R$ and continuity, $\mathbb{P}_F$ almost surely,

$$\sqrt{\frac{T_n(R)}{n}} \le \inf_{\theta \in \Theta_n \cap R} \sqrt{\int_{\mathbf{T}} \left|\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)\right|^2 d\mu(\mathbf{t}) + o(1)} \to \sqrt{\min_{\theta \in \Theta \cap R} \int_{\mathbf{T}} \left|\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)\right|^2 d\mu(\mathbf{t})}.$$

By the continuous mapping theorem, $T_n(R)/n \to \min_{\theta \in \Theta \cap R} \int_{\mathbf{T}} |\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)|^2 d\mu(\mathbf{t})$, $\mathbb{P}_F$ almost surely. □

Theorem S.1 shows that under the null hypothesis, $\Gamma_{\xi_n, F, n, R}(\mathbb{G}_{n,F})$ is, at least, a conservative approximation to $T_n(R)$ regardless how fast $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R))$ vanishes. But if it vanishes fast enough, this approximation is asymptotically exact. For example, if $|d^2 \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta_1 + \tau(\theta_2 - \theta_1))/d\tau^2|_{\tau=0}| \le C d_F(\theta_1, \theta_2)^2$, the approximation is asymptotically exact if $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = o_{\mathbb{P}_F}(n^{-1/4})$. If $\rho_{\mathbf{t}}$ is linear in θ, the approximation is asymptotically exact if $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = o_{\mathbb{P}_F}(1)$. Under a fixed alternative, $T_n(R)$ diverges to infinity at the rate n.

It is worth pointing out that the validity of the approximation does not depend on the shape of the restriction set R. In particular, R does not need to be convex. In addition, even if Assumptions B.4(i) and (ii) are not satisfied, one can still obtain a valid uniform asymptotic approximation if one replaces the term $\frac{\sqrt{n}\, d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \tilde{\theta})}{d\theta}[\theta - \tilde{\theta}]$ in $\Gamma_{\xi_n, F, n, R}(\mathbb{G}_{n,F})$ by $\sqrt{n}[\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) - \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \tilde{\theta})]$.

S.4. Proof of Theorem 5.1

To simplify notation, I suppress the dependence of Q_n^*, $\hat{\Theta}_n^*$ and θ_n^* on γ_n, λ_n, and R when there is no ambiguity. Let $c > 0$ and $\eta > 0$ be two constants. Define

$$h_{c,\eta}(x) = \begin{cases} 1 & \text{if } x > c + \eta, \\ (x - c)/\eta & \text{if } c + \eta \ge x > c, \\ 0 & \text{if } x \le c. \end{cases} \tag{S.3}$$

Lemma S.3. *Under the assumptions of Theorem* 5.1, *uniformly in* $(F, R) \in \mathcal{J}$, $T_n^*(R) \ge \Gamma_{\xi_n, F, n, R}(\mathbb{G}_n^*) + o_{\mathbb{P}_F}(1)$.

Proof. I establish this lemma by proving the following chain of inequalities:

$$\begin{aligned} T_n^*(R) &\ge \inf_{(\gamma_n, \lambda_n) \in \bar{I}_n} \inf_{\theta \in \Theta_{n,F}^{\frac{\xi_n}{2}}(R)} n Q_n^*(\theta, \gamma_n, 0) + o_{\mathbb{P}_F}(1) \ge \inf_{(\gamma_n, \lambda_n) \in \bar{I}_n} \bar{T}_n^*(R, \gamma_n) + o_{\mathbb{P}_F}(1) \\ &= \inf_{(\gamma_n, \lambda_n) \in \bar{I}_n} \tilde{T}_n^*(R, \gamma_n) + o_{\mathbb{P}_F}(1) \ge \Gamma_{\xi_n, F, n, R}\big(\mathbb{G}_n^*\big) + o_{\mathbb{P}_F}(1) \end{aligned} \tag{S.4}$$

uniformly in $(F, R) \in \mathcal{J}$, where

$$\bar{T}_n^*(R, \gamma_n) = \inf_{\tilde{\theta} \in \Theta_{n,F}(R)} \inf_{\theta \in R_{n,F}^{\frac{\xi_n}{2}}(\tilde{\theta})} \int_{\mathbf{T}} \left[\mathbb{G}_n^*(\tilde{\theta}, \mathbf{t}) + \frac{\sqrt{n}\, d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \tilde{\theta})}{d\theta}\big[\theta(\gamma_n, \theta, \tilde{\theta}) - \tilde{\theta}\big] + \Delta_n(\mathbf{t})\right]^2 d\mu(\mathbf{t}),$$

$\tilde{T}_n^*(R,\gamma_n)$

$$= \inf_{\tilde\theta\in\Theta_{n,F}(R)} \inf_{\theta\in R_{n,F}^{\frac{\xi_n}{2}}(\tilde\theta)} \int_{\mathbf{T}}\Big[\mathbb{G}_n^*(\tilde\theta,\mathbf{t}) + \frac{\sqrt{n}\,d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde\theta)}{d\theta}\big[\theta(\gamma_n,\theta,\tilde\theta)-\tilde\theta\big]\Big]^2 d\mu(\mathbf{t}),$$

$\theta(\gamma_n,\theta,\tilde\theta)=\gamma_n(\theta-\tilde\theta)+\tilde\theta$, and

$$\Delta_n(\mathbf{t}) = \gamma_n\sqrt{n}\bar\rho_{\mathbf{t}}(\theta) - \sqrt{n}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde\theta)}{d\theta}\big[\theta(\gamma_n,\theta,\tilde\theta)-\tilde\theta\big] + \mathbb{G}_n^*(\theta,\mathbf{t}) - \mathbb{G}_n^*(\tilde\theta,\mathbf{t}).$$

Again, to simplify notation, I suppress the dependence of Δ_n on F, R, θ, and $\tilde\theta$. Notice that $\bar{T}_n^*(R,\gamma_n) = \inf_{\tilde\theta\in\Theta_{n,F}(R)} \inf_{\theta\in R_{n,F}^{\frac{\xi_n}{2}}(\tilde\theta)} nQ_n^*(\theta,\gamma_n,0)$.

First, Theorem S.3 and Assumption B.5 imply that if $\theta_n^*\in\hat\Theta_n^*$, then $\theta_n^*\in\Theta_{n,F}^{\frac{\xi_n}{2}}(R)$ with probability approaching 1 uniformly in $(\gamma_n,\lambda_n)\in\bar{I}_n$ and $(F,R)\in\mathcal{J}$. Therefore, uniformly in $(F,R)\in\mathcal{J}$,

$$T_n^*(R) = \inf_{(\gamma_n,\lambda_n)\in\bar I_n} nQ_n^*\big(\theta_n^*,\gamma_n,0\big) \geq \inf_{(\gamma_n,\lambda_n)\in\bar I_n} \inf_{\theta\in\Theta_{n,F}^{\frac{\xi_n}{2}}(R)} nQ_n^*(\theta,\gamma_n,0) + o_{\mathbb{P}_F}(1).$$

This establishes the first inequality in equation (S.4). Next, notice that if $\theta\in\Theta_{n,F}^{\frac{\xi_n}{2}}(R)$, then $\theta\in R_{n,F}^{\frac{\xi_n}{2}}(\tilde\theta)$ for some $\tilde\theta\in\Theta_{n,F}(R)$. Therefore, the second inequality in equation (S.4) follows from

$$\inf_{\theta\in\Theta_{n,F}^{\frac{\xi_n}{2}}(R)} nQ_n^*(\theta,\gamma_n,0) \geq \inf_{\tilde\theta\in\Theta_{n,F}(R)} \inf_{\theta\in R_{n,F}^{\frac{\xi_n}{2}}(\tilde\theta)} nQ_n^*(\theta,\gamma_n,0) = \bar{T}_n^*(R,\gamma_n).$$

Now I prove the equality in equation (S.4). By the triangular inequality,

$$\sqrt{\inf_{(\gamma_n,\lambda_n)\in\bar I_n} \tilde{T}_n^*(R,\gamma_n)} + \delta_n \geq \sqrt{\inf_{(\gamma_n,\lambda_n)\in\bar I_n} \bar{T}_n^*(R,\gamma_n)} \geq \sqrt{\inf_{(\gamma_n,\lambda_n)\in\bar I_n} \tilde{T}_n^*(R,\gamma_n)} - \delta_n,$$

where $\delta_n = \sup_{(\gamma_n,\lambda_n)\in\bar I_n}\sup_{\tilde\theta\in\Theta_{n,F}(R)}\sup_{\theta\in R_{n,F}^{\xi_n/2}(\tilde\theta)}\|\Delta_n\|_{2,\mu}$. If δ_n is $o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$,

$$\sqrt{\inf_{(\gamma_n,\lambda_n)\in\bar I_n} \tilde{T}_n^*(R,\gamma_n)} + o_{\mathbb{P}_F}(1) = \sqrt{\inf_{(\gamma_n,\lambda_n)\in\bar I_n} \bar{T}_n^*(R,\gamma_n)}$$

uniformly in $(F,R)\in\mathcal{J}$. Then, because $\inf_{(\gamma_n,\lambda_n)\in\bar I_n}\tilde{T}_n^*(R,\gamma_n) = O_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$, Lemma S.10 implies the equality in equation (S.4). To see $\delta_n = o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$, first notice that by definition,

$$\frac{\sqrt{n}\,d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde\theta)}{d\theta}\big[\theta(\gamma_n,\theta,\tilde\theta)-\tilde\theta\big] = \frac{\sqrt{n}\gamma_n\,d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde\theta)}{d\theta}[\theta-\tilde\theta].$$

Then one can write $\Delta_n = \Delta_{n,1}+\Delta_{n,2}+\Delta_{n,3}$ where

$$\Delta_{n,1}(\mathbf{t}) = \gamma_n\sqrt{n}\big(\bar\rho_{\mathbf{t}}(\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\big),$$

$$\Delta_{n,2}(\mathbf{t}) = \gamma_n\sqrt{n}\bigg(\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta) - \frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\bigg),$$

$$\Delta_{n,3}(\mathbf{t}) = \mathbb{G}_n^*(\theta,\mathbf{t}) - \mathbb{G}_n^*(\tilde{\theta},\mathbf{t}).$$

By the triangular inequality, $\|\Delta_n\|_{2,\mu} \le \|\Delta_{n,1}\|_{2,\mu} + \|\Delta_{n,2}\|_{2,\mu} + \|\Delta_{n,3}\|_{2,\mu}$. Because $\gamma_n\sqrt{n} \le \sqrt{\kappa_n}$ and $\kappa_n \ln\ln n/n \to 0$, Definition 5.1(iii) implies that uniformly in $(F,R)\in\mathcal{J}$,

$$\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n}\sup_{\tilde{\theta}\in\Theta_{n,F}(R)}\sup_{\theta\in R_{n,F}^{\xi n/2}(\tilde{\theta})}\|\Delta_{n,1}\|_{2,\mu} \le \sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}\sqrt{\kappa_n}\big|\bar{\rho}_{\mathbf{t}}(\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\big| = o_{\mathbb{P}_F}(1).$$

Next, by the triangular inequality and the fact that $\gamma_n\sqrt{n}\le\sqrt{\kappa_n}$,

$$\|\Delta_{n,2}\|_{2,\mu} \le \sqrt{\kappa_n}\sqrt{\int_{\mathbf{T}}\bigg|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta}) - \frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\bigg|^2 d\mu(\mathbf{t})}$$
$$+ \sqrt{\kappa_n}\sqrt{\int_{\mathbf{T}}\big|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})\big|^2 d\mu(\mathbf{t})}.$$

Therefore, uniformly in $(F,R)\in\mathcal{J}$, $\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n}\sup_{\tilde{\theta}\in\Theta_{n,F}(R)}\sup_{\theta\in R_{n,F}^{\xi_n}(\tilde{\theta})}\|\Delta_{n,2}\|_{2,\mu} = o(1)$ by Assumptions B.1(ii), B.2(iii), and B.4(ii).

Lastly, if $\omega\in\ell^\infty(\Theta\times\mathbf{T})$, define $g_{F,\xi}(\omega) = \sup_{d_F(\theta,\tilde{\theta})\le\xi}\sup_{\mathbf{t}\in\mathbf{T}}|\omega(\theta,\mathbf{t}) - \omega(\tilde{\theta},\mathbf{t})|$. I first show that for any $\epsilon>0$

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}_F\big(g_{F,\xi_n}(\mathbb{G}_n^*)>\epsilon\big) = \lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{1}\big(g_{F,\xi_n}(\mathbb{G}_n^*)>\epsilon\big) = 0.$$

Notice $\mathbf{1}(g_{F,\xi_n}(\mathbb{G}_n^*)>\epsilon) \le h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi_n}(\mathbb{G}_n^*)$ where $h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}$ is defined by equation (S.3). Then $\mathcal{H} = \{h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\}$ and $\mathcal{G} = \{g_{F,\xi} : \xi>0, F\in\mathcal{F}\}$ satisfy the assumptions of Lemma S.5. Lemma S.5 implies that

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{\xi>0}\big|\mathbb{E}_F h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi}\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi}(\mathbb{G}_F)\big|$$
$$= \lim_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{\xi>0}\big|\mathbb{E}_F\mathbb{E}^* h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi}\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi}(\mathbb{G}_F)\big|$$
$$\le \lim_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{\xi>0,\tilde{F}\in\mathcal{F}}\big|\mathbb{E}_F\mathbb{E}^* h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{\tilde{F},\xi}\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{\tilde{F},\xi}(\mathbb{G}_F)\big|$$
$$\le \lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\sup_{\xi>0,\tilde{F}\in\mathcal{F}}\big|\mathbb{E}^* h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{\tilde{F},\xi}\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{\tilde{F},\xi}(\mathbb{G}_F)\big| = 0.$$

The last inequality holds by Jensen's inequality. This means that

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\big|\mathbb{E}_F h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi_n}\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi_n}(\mathbb{G}_F)\big| = 0. \tag{S.5}$$

By Definition 5.1(iii), ϱ is uniformly pre-Gaussian. Then equation (S.5) and Assumption B.4(iii) suggest

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{1}\big(g_{F,\xi_n}(\mathbb{G}_n^*)>\epsilon\big)\leq\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi_n}(\mathbb{G}_n^*)=\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}h_{\frac{\epsilon}{2},\frac{\epsilon}{2}}\circ g_{F,\xi_n}(\mathbb{G}_F)$$
$$\leq\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}\left(\sup_{d_F(\theta,\tilde\theta)\leq\xi_n}\sup_{\mathbf{t}\in\mathbf{T}}\big|\mathbb{G}_F(\theta,\mathbf{t})-\mathbb{G}_F(\tilde\theta,\mathbf{t})\big|>\frac{\epsilon}{2}\right)=0.$$

Uniformly in $(F,R)\in\mathcal{J}$, $\sup_{(\gamma_n,\lambda_n)\in\bar I_n}\sup_{\tilde\theta\in\Theta_{n,F}(R)}\sup_{\theta\in R_{n,F}^{\xi_n/2}(\tilde\theta)}\|\Delta_{n,3}\|_{2,\mu}\leq g_{F,\xi_n}(\mathbb{G}_n^*)=o_{\mathbb{P}_F}(1)$. The above bounds imply that $\delta_n=o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$ and the equality in equation (S.4) follows.

To prove the last inequality in (S.4), I only need to show that if $\theta\in R_{n,F}^{\frac{\xi_n}{2}}(\tilde\theta)$, then $\theta(\gamma_n,\theta,\tilde\theta)\in R_{n,F}^{\xi_n}(\tilde\theta)$ for sufficiently large n and all $(\gamma_n,\lambda_n)\in\bar I_n$. Notice that if n is sufficiently large, $\gamma_n\leq\sqrt{\kappa_n/n}<1$ for all $(\gamma_n,\lambda_n)\in\bar I_n$. Therefore, $\theta(\gamma_n,\theta,\tilde\theta)\in\Theta_n\cap R$ by convexity (Assumption B.2(i)) and the fact that $\theta,\tilde\theta\in\Theta_n\cap R$. Because, in addition, $d_F(\theta(\gamma_n,\theta,\tilde\theta),\tilde\theta)\leq\xi_n$, $\theta(\gamma_n,\theta,\tilde\theta)\in R_{n,F}^{\xi_n}(\tilde\theta)$ for all $(\gamma_n,\lambda_n)\in\bar I_n$ if n is sufficiently large. □

Proof of Theorem 5.1. *Proof of the first claim.* Because $T_n^*(R)$ is weakly decreasing in I_n, I only need to prove the first claim under $I_n=\bar I_n$. Notice that by Lemma S.3, $C_n^*(1-\alpha,R)\geq q_n^*(1-\alpha,\Gamma_{\xi_n,F,n,R})+o_{\mathbb{P}_F}(1)$ uniformly in $(F,R)\in\mathcal{J}$. Define $\bar{\mathcal{G}}=\{\Gamma_{\kappa,F,m,R}:\kappa\geq0,(F,R)\in\mathcal{J},m\in\mathbb{Z}^+\}$. Then by Theorem S.1,

$$\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\mathbb{P}_F\big(T_n(R)>C_n^*(1-\alpha+\eta,R)+\eta\big)$$
$$\leq\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\mathbb{P}_F\left(\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})>q_n^*\left(1-\alpha+\frac{\eta}{2},\Gamma_{\xi_n,F,n,R}\right)+\frac{\eta}{2}\right)$$
$$\leq\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\sup_{g\in\bar{\mathcal{G}}}\mathbb{P}_F\left(g(\mathbb{G}_{n,F})>q(1-\alpha,g,F)+\frac{\eta}{4}\right)$$
$$+\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\sup_{g\in\bar{\mathcal{G}}}\mathbb{P}_F\left(q(1-\alpha,g,F)>q_n^*\left(1-\alpha+\frac{\eta}{2},g\right)+\frac{\eta}{4}\right).$$

The second inequality follows because $\Gamma_{\xi_n,F,n,R}\in\bar{\mathcal{G}}$ for all n and $(F,R)\in\mathcal{J}$. Lemma S.7 ensures that $\bar{\mathcal{G}}$ satisfies the assumptions in Proposition S.1. Therefore, equations (S.6) and (S.7) of Proposition S.1 imply that the first term after the second inequality is no larger than α and the second term is 0. This implies that the test is valid if $\eta>0$.

Now suppose that in addition, Assumption B.7 holds. For any $\epsilon>0$, Proposition S.1, Lemmas S.6 and S.7 imply that for any sufficiently small $\eta>0$,

$$\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}_n^\epsilon(\alpha)}\mathbb{P}_F\big(T_n(R)>C_n^*(1-\alpha,R)\big)$$
$$\leq\limsup_{n\to\infty}\sup_{J\in\mathcal{J}_n^\epsilon(\alpha)}\mathbb{P}_F\big(\Gamma_{n,J}(\mathbb{G}_{n,F})>q(1-\alpha-\eta,\Gamma_{n,J},F)-\eta\big)$$

$$\le \limsup_{n\to\infty} \sup_{J\in\mathcal{J}_n^{\epsilon}(\alpha)} \mathbb{P}_F\big(\Gamma_{n,J}(\mathbb{G}_{n,F}) > q(1-\alpha+\eta,\Gamma_{n,J},F)+\eta\big)$$
$$+ \limsup_{n\to\infty} \sup_{J\in\mathcal{J}_n^{\epsilon}(\alpha)} \mathbb{P}_F\big(q(1-\alpha+\eta,\Gamma_{n,J},F)+\eta \ge \Gamma_{n,J}(\mathbb{G}_{n,F}) > q(1-\alpha-\eta,\Gamma_{n,J},F)-\eta\big)$$
$$\le \alpha + \limsup_{n\to\infty} \sup_{J\in\mathcal{J}_n^{\epsilon}(\alpha)} \mathbb{P}_F\big(C_F(1-\alpha+2\eta,\Gamma_{n,J})+2\eta \ge \Gamma_{n,J}(\mathbb{G}_F) > C_F(1-\alpha-2\eta,\Gamma_{n,J})-2\eta\big),$$

where $\Gamma_{n,J}=\Gamma_{\xi_n,F,n,R}$ and $J=(F,R)$. The second term in the last line converges to 0 as $\eta\to 0$ under Assumption B.7(i). This suggests that $\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}_n^{\epsilon}(\alpha)}\mathbb{P}_F(T_n(R) > C_n^*(1-\alpha,R))\le\alpha$ for all $\epsilon>0$. Therefore, there exists $\epsilon_n\downarrow 0$ such that

$$\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)\le\alpha.$$

This implies that

$$\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)$$
$$\le \limsup_{n\to\infty}\Big[\sup_{(F,R)\in\mathcal{J}\setminus\mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)$$
$$\vee \sup_{(F,R)\in\mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)\Big]$$
$$\le \Big[\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}\setminus\mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)\Big]$$
$$\vee \Big[\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F\big(T_n(R) > C_n^*(1-\alpha,R)\big)\Big].$$

Assumption B.7(ii) implies that the first term is at most α. And the second term is at most α by the above argument. Therefore, the test is valid with $\eta=0$.

Proof of the second claim. For any $(\gamma_n,\lambda_n)\in I_n$

$$\frac{T_n^*(R)}{\kappa_n} \le \frac{nQ_n^*\big(\theta_n^*,\gamma_n,0\big)}{\kappa_n} \le \frac{\inf_{\theta\in\Theta_n\cap R} nQ_n^*(\theta,\gamma_n,\lambda_n)}{\kappa_n}.$$

By Theorem 3.6.2 in Van Der Vaart and Wellner (1996), $\mathbb{G}_n^*(\theta,\mathbf{t})$ converges in distribution to $\mathbb{G}_F(\theta,\mathbf{t})$, $\mathbb{P}_F$ almost surely. This implies $\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\mathbb{G}_n^*(\theta,\mathbf{t})| = O_{\mathbb{P}^*}(1)$, $\mathbb{P}_F$ almost surely. In addition, $\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\bar\rho_{\mathbf{t}}(\theta)-\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)| = o(1)$, $\mathbb{P}_F$ almost surely. Then $\gamma_n\le\sqrt{\kappa_n/n}\to 0$ and $\kappa_n\to\infty$ imply that $\mathbb{P}_F$ almost surely

$$\frac{T_n^*(R)}{\kappa_n} \le \inf_{\theta\in\Theta_n\cap R}\Bigg\{\int_{\mathbf{T}}\big|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\big|^2 + \frac{\int_{\mathbf{T}}\mathbb{G}_n^*(\theta,\mathbf{t})^2\,d\mu(\mathbf{t})}{\kappa_n}$$
$$+ \frac{2\int_{\mathbf{T}}\big|\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta)\mathbb{G}_n^*(\theta,\mathbf{t})\big|\,d\mu(\mathbf{t})}{\sqrt{\kappa_n}} + o_{\mathbb{P}^*}(1)\Bigg\}$$
$$= \min_{\theta\in\Theta\cap R} Q_F(\theta) + o_{\mathbb{P}^*}(1).$$

Therefore, $T_n^*(R) = O_{\mathbb{P}^*}(\kappa_n)$, $\mathbb{P}_F$ almost surely. Because $\Theta \cap R$ is compact and $\Theta_F \cap R = \emptyset$, $\min_{\theta\in\Theta\cap R} Q_F(\theta) > 0$ by the continuity of $Q_F(\cdot)$. By Theorem S.1, $T_n(R)$ diverges to infinity at rate n, which is faster than κ_n. Therefore, the second claim holds. If $(0, \kappa_n) \in I_n$, then $T_n^*(R) \le \int_{\mathbf{T}} |\mathbb{G}_n^*(\theta_n^*, \mathbf{t})|^2 \, d\mu = O_{\mathbb{P}^*}(1)$, $\mathbb{P}_F$ almost surely. Hence, under a fixed alternative, the bootstrap critical value does not diverge. □

S.5. Power comparison with CNS: An example

In this section, I show that one can construct examples where for each sequence of rescaling parameters in my test, there exists a sequence of tuning parameters in CNS and a sequence of data generating processes along which the test based on CNS has better size control or power. Similarly, for each sequence of tuning parameters in CNS, one can find a sequence of rescaling parameters and a sequence of data generating processes along which my test has better size control or power.

Example S.1 (Asymptotic Size). Consider the case where $W_i = (X_i, Y_i)$ is bivariate normal with the identity covariance matrix and has mean $\theta_{F_n} = (\theta_{1,n}, \theta_{2,n})$ as $n \to \infty$. One would like to test $H_0 : \theta_F \ge 0$ against $H_1 : \theta_F \ngeq 0$. Then the moment conditions are $\mathbb{E}_F(W_i - \theta) = 0$, where $\theta = (\theta_1, \theta_2)$. The test statistic that gives equal weights to both moment conditions is

$$T_n(R) = \min_{(\theta_1,\theta_2)\ge 0} \frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (\bar{X} - \theta_{1,n}) - \sqrt{n}(\theta_1 - \theta_{1,n})\right]^2 + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (\bar{Y} - \theta_{2,n}) - \sqrt{n}(\theta_2 - \theta_{2,n})\right]^2}{2}$$

$$= \min_{(h_1,h_2)\in\mathbf{V}_n(\theta_{F_n})} \frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (\bar{X} - \theta_{1,n}) - h_1\right]^2 + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (\bar{X} - \theta_{1,n}) - h_2\right]^2}{2},$$

where the true local parameter space is

$$\mathbf{V}_n(\theta_{F_n}) = \left\{(h_1, h_2) : h_1/\sqrt{n} \ge -\theta_{1,n},\, h_2/\sqrt{n} \ge -\theta_{2,n}\right\}.$$

Similarly, the bootstrap statistic with $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$ is

$$T_n^*(R) = \min_{(h_1,h_2)\in\mathbf{V}_n^{\gamma_n}(\theta_{F_n})} \frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \big(X_i^* - \bar{X}\big) - h_1\right]^2 + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \big(Y_i^* - \bar{Y}\big) - h_2\right]^2}{2} + o_{\mathbb{P}_{F_n}}(1),$$

where $\mathbf{V}_n^{\gamma_n}(\theta_{F_n}) = \{(h_1, h_2) : h_1/\sqrt{n} \ge -\gamma_n\theta_{1,n}, h_2/\sqrt{n} \ge -\gamma_n\theta_{2,n}\}$.

If one follows the method proposed in CNS to estimate the local parameter space, one can obtain a bootstrap statistic

$$T_{n,\mathrm{CNS}}^*(R) = \min_{(h_1,h_2)\in\tilde{\mathbf{V}}_n^{r_n}(\hat{\theta}_n)} \frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \big(X_i^* - \bar{X}\big) - h_1\right]^2 + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \big(Y_i^* - \bar{Y}\big) - h_2\right]^2}{2},$$

where $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ is the minimizer obtained from the minimization problem for $T_n(R)$, r_n is a sequence of positive numbers that converges to 0, and

$$\tilde{\mathbf{V}}_n^{r_n}(\hat{\theta}_n) = \left\{(h_1, h_2) : \frac{h_1}{\sqrt{n}} \geq -\hat{\theta}_{1,n} + \min(\hat{\theta}_{1,n}, r_n), \frac{h_2}{\sqrt{n}} \geq -\hat{\theta}_{2,n} + \min(\hat{\theta}_{2,n}, r_n)\right\}.$$

Given $\gamma_n = O(\sqrt{1/\ln n})$, let $r_n = 1/\gamma_n\sqrt{n}$, $\theta_{1,n} = 0$ and $\theta_{2,n} = c/\gamma_n\sqrt{n}$ where c is a constant. And let Z_1 and Z_2 be two independent standard normal random variables. Along F_n, $T_n(R)$ and $T_n^*(R)$ converge to $\min(Z_1, 0)^2/2$ and $[\min(Z_1, 0)^2 + \min(Z_2 + c, 0)^2]/2$ in distribution, respectively.

Suppose $c = 1.01$. Because $\hat{\theta}_{2,n} = \theta_{2,n} + O_{\mathbb{P}_{F_n}}(n^{-1/2})$,

$$\mathbb{P}_{F_n}(\hat{\theta}_{2,n} > r_n) = \mathbb{P}_{F_n}(\hat{\theta}_{2,n} - \theta_{2,n} > r_n - \theta_{2,n}) = \mathbb{P}_{F_n}\left(O_{\mathbb{P}_{F_n}}(n^{-1/2}) > \frac{1-c}{\gamma_n\sqrt{n}}\right)$$

converges to 1. This suggests that along F_n, $T_{n,\mathrm{CNS}}^*(R)$ converges to $\min(Z_1, 0)^2/2$. Therefore, $T_{n,\mathrm{CNS}}^*(R)$ leads to a test that has the exact asymptotic size for all $\alpha > 0.5$ along F_n. On the other hand, a 10% test using $T_n^*(R)$ has a rejection probability that converges to 8.75%, while a 5% test has a rejection probability that converges to 4.4%. Therefore, $T_{n,\mathrm{CNS}}^*(R)$ has better size control.

Now suppose $c = 0.99$. Given any r_n such that $\sqrt{\ln n/n} = O(r_n)$, let $\gamma_n = 1/r_n\sqrt{n}$. Let $\theta_{1,n}$ and $\theta_{2,n}$ be the same as the above. Along F_n, $\mathbb{P}_{F_n}(\hat{\theta}_{2,n} > r_n) \to 0$ and $T_{n,\mathrm{CNS}}^*(R)$ converges to $[\min(Z_1, 0)^2 + \min(Z_2, 0)^2]/2$. Then a 10% test and a 5% test based on $T_{n,\mathrm{CNS}}^*(R)$ have asymptotic sizes 4.29% and 1.99%, respectively. The corresponding values based on $T^*(R)$ are 8.69% and 4.37%, respectively. Along this F_n, my method has better size control.

Example S.2 (Local Power). Consider the same testing problem as in the previous example. Construct F_n in the same way as above except that $\theta_{1,n} = -1/\sqrt{n}$. Then F_n is a sequence of local alternatives. Along F_n, the limiting distribution of $T_n(R)$ is $\min(Z_1 - 1, 0)^2/2$. The limiting distributions of the bootstrap statistics are the same as the ones calculated in the previous example. If $c = 1.01$, a 5% test using $T_n^*(R)$ rejects with a probability converging to 36.07%, while a 5% test using $T_{n,\mathrm{CNS}}^*(R)$ rejects with a probability converging to 38.91%, which is about 3% higher. If $c = 0.99$, a 5% test based on $T_n^*(R)$ has a rejection probability that converges to 35.94%, while a test based on $T_{n,\mathrm{CNS}}^*(R)$ has a rejection probability that converges to 23.63%. Now $T_n^*(R)$ has better power. A similar result holds for a 10% test.

S.6. Useful results

Lemma S.4. *If Assumptions* B.1–B.3 *hold, then* $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = o_{\mathbb{P}_F}(1)$ *uniformly in* $(F, R) \in \mathcal{J}$.

Theorem S.2. *If Assumptions* B.1 *to* B.3 *hold, then uniformly in* $(F, R) \in \mathcal{J}$, $\vec{d}_{w,F}(\hat{\Theta}_n(R), \Theta_F \cap R) = O_{\mathbb{P}_F}(n^{-1/2})$, $\vec{d}_{w,F}(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = O_{\mathbb{P}_F}(n^{-1/2})$, *and* $\vec{d}_F(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = O_{\mathbb{P}_F}(\zeta_n^{\frac{1}{\nu}} n^{-\frac{1}{2\nu}})$.

The first claim in Theorem S.2 says that under $\vec{d}_{w,F}$, the convergence rate is parametric even though θ contains unknown functions. Hong (2017) obtains the same convergence rates in conditional moment equality models with semi/nonparametric unknown parameters. My result is stronger in the sense that the rates are valid uniformly on $\mathcal{J}$. It holds because Definition 5.1(iii) implicitly restricts the complexity of Θ. With a more complex parameter space, the convergence rate can be lower. The convergence rate under d_F is generally slower than the parametric rate because $\zeta_n \to \infty$.

THEOREM S.3. *If Assumptions* B.1 *to* B.3 *hold and* $\kappa_n \to \infty$, $\kappa_n \ln\ln n/n \to 0$, *then*

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\mathbb{P}_F\Big(\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n}\vec{d}_{w,F}\big(\hat{\Theta}_n^*(\gamma_n,\lambda_n,R),\Theta_F\cap R\big)>M\kappa_n^{-\frac{1}{2}}\Big)=0,$$

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\mathbb{P}_F\Big(\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n}\vec{d}_{w,F}\big(\hat{\Theta}_n^*(\gamma_n,\lambda_n,R),\Theta_{n,F}(R)\big)>M\kappa_n^{-\frac{1}{2}}\Big)=0,$$

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{(F,R)\in\mathcal{J}}\mathbb{P}_F\Big(\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n}\vec{d}_{F}\big(\hat{\Theta}_n^*(\gamma_n,\lambda_n,R),\Theta_{n,F}(R)\big)>M\zeta_n^{\frac{1}{\nu}}\kappa_n^{-\frac{1}{2\nu}}\Big)=0.$$

Notice that the convergence rate of $\hat{\Theta}_n^*(\gamma_n,\lambda_n,R)$ has an upper bound that depends only on κ_n, which allows me to freely change γ_n without significantly changing the imposed identified set. This is possible because of the additional penalty term $\lambda_n Q_n(\theta)$.

PROPOSITION S.1. *Let* $\mathbf{M}(\cdot,\cdot)$ *be a function on* $\mathbb{R}^2$ *such that* $\sup_{(x_1,x_2)\in\mathcal{K}}\mathbf{M}(x_1,x_2)<\infty$ *for any bounded set* $\mathcal{K}\subset\mathbb{R}^2$. $\mathcal{G}$ *is a collection of nonnegative functions on* $\ell^\infty(\Theta\times\mathbf{T})$ *such that for any* $g\in\mathcal{G}$ *and* $\omega_1,\omega_2\in\ell^\infty(\Theta\times\mathbf{T})$,

$$\big|g(\omega_1)-g(\omega_2)\big|<\mathbf{M}\big(\|\omega_1\|_\infty,\|\omega_2\|_\infty\big)d_\infty(\omega_1,\omega_2).$$

Then, for any $\eta>0$ *and* $\alpha\in[0,1]$

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\mathbb{P}_F\big(g(\mathbb{G}_{n,F})>q(1-\alpha,g,F)+\eta\big)\le\alpha, \tag{S.6}$$

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\mathbb{P}_F\big(q(1-\alpha,g,F)>q_n^*(1-\alpha+\eta,g)+\eta\big)=0. \tag{S.7}$$

LEMMA S.5. *Let* $\mathcal{H}$ *be a class of bounded Lipschitz functions with Lipschitz constant* γ *and bound* $\mathbf{B}$. *If the set of functions* $\mathcal{G}$ *satisfies the assumptions of Proposition* S.1, *then*

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\sup_{h\in\mathcal{H}}\sup_{g\in\mathcal{G}}\big|\mathbb{E}^*h\circ g\big(\mathbb{G}_n^*\big)-\mathbb{E}h\circ g(\mathbb{G}_F)\big|=0, \tag{S.8}$$

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{h\in\mathcal{H}}\sup_{g\in\mathcal{G}}\big|\mathbb{E}_F h\circ g(\mathbb{G}_{n,F})-\mathbb{E}h\circ g(\mathbb{G}_F)\big|=0. \tag{S.9}$$

LEMMA S.6. *Under the assumptions of Proposition* S.1, *for any* $\eta>0$,

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\sup_{a\le b\in\mathbb{R}}\big[\mathbb{P}_F\big(g(\mathbb{G}_{n,F})\in[a,b]\big)-\mathbb{P}\big(g(\mathbb{G}_F)\in[a-\eta,b+\eta]\big)\big]\le 0.$$

LEMMA S.7. *For all* $m \in \mathbb{Z}^+$, $(F,R) \in \mathcal{J}$, $\kappa \geq 0$, *and* $\omega_1, \omega_2 \in \ell^\infty(\Theta \times \mathbf{T})$,

$$\left|\Gamma_{\kappa,F,m,R}(\omega_1) - \Gamma_{\kappa,F,m,R}(\omega_2)\right| \leq 4\max\left(\|\omega_1\|_\infty, \|\omega_2\|_\infty\right) d_\infty(\omega_1, \omega_2).$$

S.7. PROOFS FOR RESULTS IN SECTION S.6

PROOF OF LEMMA S.4. Notice that Lemmas S.8 and S.9 imply that $\forall \epsilon > 0$, uniformly in $(F,R) \in \mathcal{J}$,

$$\begin{aligned}
&\mathbb{P}_F\left(\vec{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) > \epsilon\right) \\
&\leq \mathbb{P}_F\left(\inf_{\theta \in \Theta_n \cap R : d_F(\theta, \Theta_{n,F}(R)) > \epsilon} Q_n(\theta) \leq \inf_{\Theta_n \cap R} Q_n(\theta)\right) \\
&\leq \mathbb{P}_F\left(\inf_{\theta \in \Theta_n \cap R : d_F(\theta, \Theta_{n,F}(R)) > \epsilon} Q_F(\theta) - \inf_{\Theta_n \cap R} Q_F(\theta) \leq O_{\mathbb{P}_F}\left(\frac{1}{\sqrt{n}}\right)\right) \\
&\leq \mathbb{P}_F\left(S_n(\epsilon) \leq O_{\mathbb{P}_F}\left(\frac{1}{\sqrt{n}}\right)\right) \to 0.
\end{aligned}$$

The third inequality holds by Assumption B.3(i). □

PROOF OF THEOREM S.2. I begin with the first claim. Let $\theta_{F,R}$ be a point in $\Theta_F \cap R$. This point exists as long as $(F,R) \in \mathcal{J}$. By Lemma S.9,

$$\left|Q_n(\Pi_{n,R}\theta_{F,R}) - Q_F(\Pi_{n,R}\theta_{F,R})\right| \leq 2d_{w,F}(\Pi_{n,R}\theta_{F,R}, \theta_{F,R})A_{n,F} + B_{n,F},$$

where $A_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})$, $B_{n,F} = O_{\mathbb{P}_F}(n^{-1})$ uniformly in $F \in \mathcal{F}$. By Assumptions B.1(ii) and B.2(iii),

$$2d_{w,F}(\Pi_{n,R}\theta_{F,R}, \theta_{F,R})A_{n,F} \leq \frac{2C_1}{C_2} d_s(\Pi_{n,R}\theta_{F,R}, \theta_{F,R})A_{n,F} = o_{\mathbb{P}_F}\left(\frac{1}{n}\right),$$

which implies uniformly in $(F,R) \in \mathcal{J}$,

$$\left|Q_n(\Pi_{n,R}\theta_{F,R}) - Q_F(\Pi_{n,R}\theta_{F,R})\right| = O_{\mathbb{P}_F}\left(\frac{1}{n}\right). \tag{S.10}$$

Because $d_{w,F}(\theta, \Theta_F) = d_{w,F}(\theta, \Theta_F \cap R)$ for $(F,R) \in \mathcal{J}$, Lemma S.9 implies

$$Q_n(\theta) \geq Q_F(\theta) - 2d_{w,F}(\theta, \Theta_F \cap R)A_{n,F} - B_{n,F}. \tag{S.11}$$

Therefore, for sufficiently large n

$$\begin{aligned}
&\mathbb{P}_F\left(\sqrt{n}\vec{d}_{w,F}\left(\hat{\Theta}_n(R), \Theta_F \cap R\right) > M\right) \\
&\quad \leq \mathbb{P}_F\left(\inf_{\theta \in \Theta_n \cap R : d_{w,F}(\theta, \Theta_F \cap R) > \frac{M}{\sqrt{n}}} Q_n(\theta) \leq Q_n(\Pi_{n,R}\theta_{F,R})\right) \\
&\quad \leq \mathbb{P}_F\left(\inf_{\theta \in \Theta_n \cap R : d_{w,F}(\theta, \Theta_F \cap R) > \frac{M}{\sqrt{n}}} \left[Q_F(\theta) - 2d_{w,F}(\theta, \Theta_F \cap R)A_{n,F} - B_{n,F}\right] \leq Q_F(\Pi_{n,R}\theta_{F,R}) + O_{\mathbb{P}_F}(n^{-1})\right) \\
&\quad \leq \mathbb{P}_F\left(\inf_{x \geq M}\left(x^2 - 2x\sqrt{n}A_{n,F}\right) \leq O_{\mathbb{P}_F}(1)\right).
\end{aligned} \tag{S.12}$$

The first inequality holds by Lemma S.8 and the fact that $\Pi_{n,R}\theta_{F,R} \in \Theta_n \cap R$. The second inequality holds by equations (S.10) and (S.11). Because $Q_F(\theta) = d_{w,F}(\theta, \Theta_F \cap R)^2$ and $B_{n,F} = O_{\mathbb{P}_F}(n^{-1})$, one can multiply both sides by n and use the change of variable $x = \sqrt{n} d_{w,F}(\theta, \Theta_F \cap R)$ to obtain the last inequality. Lastly, notice that $\sqrt{n}A_{n,F} = O_{\mathbb{P}_F}(1)$ uniformly in $(F, R) \in \mathcal{J}$. As $M \to \infty$, the probability in the last line of equation (S.12) converges to 0 uniformly in $(F, R) \in \mathcal{J}$ because $\inf_{x \geq M}(x^2 - 2x\sqrt{n}A_{n,F})$ diverges to infinity. This concludes the first claim.

For the second claim, notice that for any $\theta_1 \in \Theta$, $\theta_2 \in \Theta_F \cap R$, and $\theta_3 \in \Theta_{n,F}(R)$,

$$d_{w,F}(\theta_1, \theta_2) \leq d_{w,F}(\theta_1, \theta_3) + d_{w,F}(\theta_2, \theta_3).$$

By the definition of $\Theta_{n,F}(R)$,

$$d_{w,F}(\theta_2, \theta_3) = \sqrt{Q_F(\theta_3)} \leq \sqrt{Q_F(\Pi_{n,R}\theta_2)} = d_{w,F}(\Pi_{n,R}\theta_2, \theta_2) = o\big(n^{-1/2}\big).$$

This shows that $d_{w,F}(\theta_1, \theta_2) \leq d_{w,F}(\theta_1, \theta_3) + o(n^{-1/2})$. Because this holds for all $\theta_2 \in \Theta_F \cap R$ and $\theta_3 \in \Theta_{n,F}(R)$, $d_{w,F}(\theta_1, \Theta_F \cap R) \leq d_{w,F}(\theta_1, \Theta_{n,F}(R)) + o(n^{-1/2})$ uniformly in $(F, R) \in \mathcal{J}$. Similarly, $d_{w,F}(\theta_1, \Theta_{n,F}(R)) \leq d_{w,F}(\theta_1, \Theta_F \cap R) + o(n^{-1/2})$. This implies that uniformly in $\Theta_1 \subset \Theta$ and $(F, R) \in \mathcal{J}$,

$$\vec{d}_{w,F}\big(\Theta_1, \Theta_{n,F}(R)\big) = \vec{d}_{w,F}(\Theta_1, \Theta_F \cap R) + o\big(n^{-1/2}\big). \tag{S.13}$$

Therefore, $\vec{d}_{w,F}(\hat{\Theta}_n(R), \Theta_{n,F}(R)) = \vec{d}_{w,F}(\hat{\Theta}_n(R), \Theta_F \cap R) + o(n^{-1/2}) = O_{\mathbb{P}_F}(1/\sqrt{n})$.

For the last claim, by Lemma S.4, $\hat{\Theta}_n(R) \subset \Theta^{\epsilon}_{n,F}(R)$ with probability converging to 1 uniformly on $\mathcal{J}$. Then by Assumption B.3(ii),

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\left(\vec{d}_F\big(\hat{\Theta}_n(R), \Theta_{n,F}(R)\big) > M\left(\frac{\zeta_n}{\sqrt{n}}\right)^{\frac{1}{\nu}}\right) \\
&\leq \limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\left(\sup_{\theta\in\hat{\Theta}_n(R)} d_F\big(\theta, \Theta_{n,F}(R)\big)^{\nu} > M^{\nu}\frac{\zeta_n}{\sqrt{n}}\right) \\
&\leq \limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\Big(\sqrt{n}\zeta_n \sup_{\theta\in\hat{\Theta}_n(R)} d_{w,F}\big(\theta, \Theta_{n,F}(R)\big) > M^{\nu}\zeta_n\Big).
\end{aligned}$$

As M goes to infinity, the last term goes to zero by the second part of this proposition. □

Proof of Theorem S.3. To simplify exposition, I suppress the dependence of $\hat{\Theta}^*_n$ on γ_n, λ_n, R and the dependence of Q^*_n on γ_n, λ_n. Again, $\theta_{F,R}$ is an arbitrary point in $\Theta_F \cap R$. Because $\sqrt{n}\gamma_n \leq \sqrt{\kappa_n}$,

$$\begin{aligned}
nQ^*_n(\theta) &= \int_{\mathbf{T}} \mathbb{G}^*_n(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t}) + 2\sqrt{n}\gamma_n \int_{\mathbf{T}} \mathbb{G}^*_n(\theta, \mathbf{t})\bar{\rho}_{\mathbf{t}}(\theta) \, d\mu(\mathbf{t}) + \kappa_n \int_{\mathbf{T}} \bar{\rho}_{\mathbf{t}}(\theta)^2 \, d\mu(\mathbf{t}) \\
&\geq \int_{\mathbf{T}} \mathbb{G}^*_n(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t}) + A_n(\theta) - B_n(\theta) \equiv Q^*_{n,1}(\theta), \\
nQ^*_n(\Pi_{n,R}\theta_{F,R}) &\leq \int_{\mathbf{T}} \mathbb{G}^*_n(\Pi_{n,R}\theta_{F,R}, \mathbf{t})^2 \, d\mu(\mathbf{t}) + A_n(\Pi_{n,R}\theta_{F,R}) + B_n(\Pi_{n,R}\theta_{F,R}) \\
&\equiv Q^*_{n,2}(\Pi_{n,R}\theta_{F,R}),
\end{aligned}$$

where

$$A_n(\theta) = \kappa_n \int_{\mathbf{T}} \bar{\rho}_{\mathbf{t}}(\theta)^2 \, d\mu(\mathbf{t}), \qquad B_n(\theta) = 2\sqrt{\kappa_n \int_{\mathbf{T}} \mathbb{G}_n^*(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t}) \int_{\mathbf{T}} \bar{\rho}_{\mathbf{t}}(\theta)^2 \, d\mu(\mathbf{t})}.$$

Notice $Q_{n,1}^*$ and $Q_{n,2}^*$ are both independent of (γ_n, λ_n). By Lemma S.8 and the fact that $\Pi_{n,R}\theta_{F,R} \in \Theta_n \cap R$,

$$\begin{aligned}
&\mathbb{P}_F\Big(\sup_{(\gamma_n,\lambda_n)\in\bar{I}_n} \sqrt{\kappa_n}\vec{d}_{w,F}\big(\hat{\Theta}_n^*, \Theta_F \cap R\big) > M\Big) \\
&\quad \le \mathbb{P}_F\Big(\exists(\gamma_n, \lambda_n) \in \bar{I}_n : \inf_{\theta\in\Theta\cap R: d_{w,F}(\theta,\Theta_F\cap R) > \frac{M}{\sqrt{\kappa_n}}} nQ_n^*(\theta) \le nQ_n^*(\Pi_{n,R}\theta_{F,R})\Big) \\
&\quad \le \mathbb{P}_F\Big(\inf_{\theta\in\Theta\cap R: d_{w,F}(\theta,\Theta_F\cap R) > \frac{M}{\sqrt{\kappa_n}}} \big[Q_{n,1}^*(\theta) - Q_{n,2}^*(\Pi_{n,R}\theta_{F,R})\big] \le 0\Big).
\end{aligned}$$

To prove the first claim, I only need to show that for M sufficiently large, the last line is arbitrarily small uniformly in $(F, R) \in \mathcal{J}$ as $n \to \infty$.

To see this, first notice by the triangular inequality,

$$\sqrt{\kappa_n Q_F(\theta)} + \sqrt{\frac{\kappa_n}{n}} \sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} \big|\mathbb{G}_{n,F}(\theta, \mathbf{t})\big| \ge \sqrt{A_n(\theta)} \ge \sqrt{\kappa_n Q_F(\theta)} - \sqrt{\frac{\kappa_n}{n}} \sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} \big|\mathbb{G}_{n,F}(\theta, \mathbf{t})\big|.$$

Because $\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} |\mathbb{G}_{n,F}(\theta, \mathbf{t})| = O_{\mathbb{P}_F}(1)$ uniformly in $F \in \mathcal{F}$ and $\kappa_n/n = o(1)$, this implies $\forall\theta \in R$

$$\sqrt{A_n(\theta)} = \sqrt{\kappa_n Q_F(\theta)} + o_{\mathbb{P}_F}(1) = \sqrt{\kappa_n} d_{w,F}(\theta, \Theta_F \cap R) + o_{\mathbb{P}_F}(1) \tag{S.14}$$

uniformly in $F \in \mathcal{F}$. Then by Assumption B.2(iii), uniformly in $(F, R) \in \mathcal{J}$,

$$\sqrt{A_n(\Pi_{n,R}\theta_{F,R})} = o\left(\sqrt{\frac{\kappa_n}{n}}\right) + o_{\mathbb{P}_F}(1) = o_{\mathbb{P}_F}(1). \tag{S.15}$$

Because ϱ is uniformly Donsker, by Lemma A.2. in Linton, Song, and Whang (2010), for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{P}_F\Big(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} \big|\mathbb{G}_n^*(\theta, \mathbf{t})\big| > M\Big) \le \sup_{F\in\mathcal{F}} \mathbb{P}_F\Big(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} \big|\mathbb{G}_F(\theta, \mathbf{t})\big| > M - \epsilon\Big),$$

which converges to 0 as $M \to \infty$. Hence, $\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} |\mathbb{G}_n^*(\theta, \mathbf{t})| = O_{\mathbb{P}_F}(1)$ uniformly in $F \in \mathcal{F}$. This and equation (S.15) imply that

$$\sqrt{B_n(\Pi_{n,R}\theta_{F,R})} = 2\sqrt{\int_{\mathbf{T}} \mathbb{G}_n^*(\Pi_{n,R}\theta_{F,R}, \mathbf{t})^2 \, d\mu(\mathbf{t})}\sqrt{A_n(\Pi_{n,R}\theta_{F,R})} = o_{\mathbb{P}_F}(1)$$

uniformly in $(F, R) \in \mathcal{J}$.

Next, notice by equation (S.14), $\forall \theta \in R$,

$$A_n(\theta) - B_n(\theta) = \left[\sqrt{\kappa_n} d_{w,F}(\theta, \Theta_F \cap R) + o_{\mathbb{P}_F}(1)\right] \times \left(\sqrt{\kappa_n} d_{w,F}(\theta, \Theta_F \cap R) + o_{\mathbb{P}_F}(1) - 2\sqrt{\int_{\mathbf{T}} \mathbb{G}_n^*(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t})}\right).$$

Define $x = \sqrt{\kappa_n} d_{w,F}(\theta, \Theta_F \cap R)$. Then $\forall \theta \in R$, uniformly in $(F, R) \in \mathcal{J}$,

$$\begin{aligned} Q_{n,1}^*(\theta) - Q_{n,2}^*(\Pi_{n,R}\theta_{F,R}) &= \int_{\mathbf{T}} \mathbb{G}_n^*(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t}) - \int_{\mathbf{T}} \mathbb{G}_n^*(\Pi_{n,R}\theta_{F,R}, \mathbf{t})^2 \, d\mu(\mathbf{t}) \\ &\quad + A_n(\theta) - B_n(\theta) - \left[A_n(\Pi_{n,R}\theta_{F,R}) + B_n(\Pi_{n,R}\theta_{F,R})\right] \\ &= -\left|O_{\mathbb{P}_F}(1)\right| + x^2 - \left|O_{\mathbb{P}_F}(1)\right| x + o_{\mathbb{P}_F}(1). \end{aligned}$$

Then the first claim of the lemma follows because

$$\begin{aligned} &\limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\left(\inf_{\theta\in\Theta\cap R: d_{w,F}(\theta,\Theta_F\cap R) > \frac{M}{\sqrt{\kappa_n}}} \left[Q_{n,1}^*(\theta) - Q_{n,2}^*(\Pi_{k_n,R}\theta_{F,R})\right] \le 0\right) \\ &= \limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{(F,R)\in\mathcal{J}} \mathbb{P}_F\left(\inf_{x>M}\left(-\left|O_{\mathbb{P}_F}(1)\right| + x^2 - \left|O_{\mathbb{P}_F}(1)\right| x + o_{\mathbb{P}_F}(1)\right) \le 0\right) = 0. \end{aligned}$$

The second claim holds because of equation (S.13). The third claim follows by Assumption B.3(ii). □

Proof of Proposition S.1. I start with proving equation (S.6). Let $h_{c,\eta}$ be defined as in equation (S.3). Then $\mathcal{H} = \{h_{c,\eta} : c \in \mathbb{R}\}$ is a collection of Lipschitz functions with Lipschitz constant $1/\eta$ and upper bound 1. By equation (S.9) in Lemma S.5,

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \sup_{c\in\mathbb{R}} \left|\mathbb{E}_F h_{c,\eta} \circ g(\mathbb{G}_{n,F}) - \mathbb{E} h_{c,\eta} \circ g(\mathbb{G}_F)\right| = 0. \tag{S.16}$$

Because $\mathbb{E}_F h_{c,\eta} \circ g(\mathbb{G}_{n,F}) \ge \mathbb{P}_F(g(\mathbb{G}_{n,F}) > c + \eta)$ and $\mathbb{E}_F h_{c,\eta} \circ g(\mathbb{G}_F) \le \mathbb{P}(g(\mathbb{G}_F) > c)$, equation (S.16) implies

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \sup_{c\in\mathbb{R}} \left[\mathbb{P}_F\left(g(\mathbb{G}_{n,F}) > c + \eta\right) - \mathbb{P}\left(g(\mathbb{G}_F) > c\right)\right] \le 0.$$

Therefore,

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \left[\mathbb{P}_F\left(g(\mathbb{G}_{n,F}) > q(1-\alpha, g, F) + \eta\right) - \mathbb{P}\left(g(\mathbb{G}_F) > q(1-\alpha, g, F)\right)\right] \\ &\le \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \sup_{c\in\mathbb{R}} \left[\mathbb{P}_F\left(g(\mathbb{G}_{n,F}) > c + \eta\right) - \mathbb{P}\left(g(\mathbb{G}_F) > c\right)\right] \le 0. \end{aligned}$$

This implies equation (S.6) because $\mathbb{P}(g(\mathbb{G}_F) > q(1-\alpha, g, F)) \le \alpha$.

Now I show equation (S.7). Because $h_{q(1-\alpha,g,F)-\eta,\eta}(x) < \mathbf{1}(x > q(1-\alpha,g,F)-\eta)$ and $\mathbb{E}h_{q(1-\alpha,g,F)-\eta,\eta} \circ g(\mathbb{G}_F) \geq \alpha$, $\forall \epsilon > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \mathbb{P}_F\big(q(1-\alpha,g,F) > q_n^*(1-\alpha+\epsilon,g)+\eta\big) \\
&\quad \leq \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \mathbb{P}_F\big(\mathbb{E}^*\mathbf{1}\big(g\big(\mathbb{G}_n^*\big) > q(1-\alpha,g,F)-\eta\big) < \alpha-\epsilon\big) \\
&\quad \leq \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \mathbb{P}_F\big(\mathbb{E}^* h_{q(1-\alpha,g,F)-\eta,\eta} \circ g\big(\mathbb{G}_n^*\big) - \alpha < -\epsilon\big) \\
&\quad \leq \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{g\in\mathcal{G}} \mathbb{P}_F\big(\mathbb{E}^* h_{q(1-\alpha,g,F)-\eta,\eta} \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{q(1-\alpha,g,F)-\eta,\eta} \circ g(\mathbb{G}_F) < -\epsilon\big) \\
&\quad \leq \frac{1}{\epsilon} \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h_{q(1-\alpha,g,F)-\eta,\eta} \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{q(1-\alpha,g,F)-\eta,\eta} \circ g(\mathbb{G}_F)\big|.
\end{aligned}$$

Notice that $\mathcal{G}$ and $\mathcal{H}$ satisfy the assumptions of Lemma S.5. Then equation (S.7) follows because by equation (S.8),

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h_{q(1-\alpha,g,F),\eta} \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{q(1-\alpha,g,F),\eta} \circ g(\mathbb{G}_F)\big| \\
&\quad \leq \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F \sup_{c\in\mathbb{R}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h_{c,\eta} \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}h_{c,\eta} \circ g(\mathbb{G}_F)\big| = 0. \qquad \square
\end{aligned}$$

Proof of Lemma S.5. I only prove equation (S.8). Equation (S.9) follows in a similar way. First, let K be a positive number. For $g \in \mathcal{G}$, define $g_K(\omega) = g(\omega_K)$ where $\omega_K(\theta,\mathbf{t}) = \operatorname{sign}(\omega(\theta,\mathbf{t}))\min(|\omega(\theta,t)|,K)$ and $\mathcal{G}_K = \{g_K : g \in \mathcal{G}\}$. Notice $\omega_K(\theta,\mathbf{t}) \in \ell^\infty(\Theta \times \mathbf{T})$, and it is bounded by K. Therefore, $h \circ g_K(\omega)$ is Lipschitz with a constant $\sup_{|x_1|\leq K,|x_2|\leq K} \mathbf{M}(x_1,x_2)\gamma$ and has a bound $\mathbf{B}$ for all $g \in \mathcal{G}$ and $h \in \mathcal{H}$. By the triangular inequality,

$$\begin{aligned}
&\mathbb{E}_F \sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}h \circ g(\mathbb{G}_F)\big| \leq D_1 + D_2 + D_3, \\
&D_1 = \mathbb{E}_F \sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h \circ g\big(\mathbb{G}_n^*\big) - \mathbb{E}^* h \circ g_K\big(\mathbb{G}_n^*\big)\big|, \\
&D_2 = \mathbb{E}_F \sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}h \circ g(\mathbb{G}_F) - \mathbb{E}h \circ g_K(\mathbb{G}_F)\big|, \\
&D_3 = \mathbb{E}_F \sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h \circ g_K\big(\mathbb{G}_n^*\big) - \mathbb{E}h \circ g_K(\mathbb{G}_F)\big|.
\end{aligned}$$

Notice $D_1 \leq 2\mathbf{B}\mathbb{P}_F(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} |\mathbb{G}_n^*(\theta,\mathbf{t})| > K)$ and $D_2 \leq 2\mathbf{B}\mathbb{P}(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}} |\mathbb{G}_F(\theta,\mathbf{t})| > K)$. Lastly, because h is bounded by $\mathbf{B}$, $\forall \epsilon > 0$,

$$D_3 \leq \epsilon + 2\mathbf{B} \sup_{F\in\mathcal{F}} \mathbb{P}_F\Big(\sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} \big|\mathbb{E}^* h \circ g_K\big(\mathbb{G}_n^*\big) - \mathbb{E}h \circ g_K(\mathbb{G}_F)\big| > \epsilon\Big).$$

By Lemma A.2 in Linton, Song, and Whang (2010), the second term on the right-hand side converges to 0 as $n \to \infty$. Hence, $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} D_3 \leq \epsilon$ for every $\epsilon > 0$. Consequently, $\forall\epsilon>0, K>0$,

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F\sup_{h\in\mathcal{H}}\sup_{g\in\mathcal{G}}\left|\mathbb{E}^*h\circ g\big(\mathbb{G}_n^*\big)-\mathbb{E}h\circ g(\mathbb{G}_F)\right|\\
&\quad\le 2\mathbf{B}\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}_F\Big(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}\left|\mathbb{G}_n^*(\theta,\mathbf{t})\right|>K\Big)\\
&\qquad+2\mathbf{B}\lim_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{P}\Big(\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}\left|\mathbb{G}_F(\theta,\mathbf{t})\right|>K\Big)+\epsilon.
\end{aligned}$$

By Lemma A.2. in Linton, Song, and Whang (2010) and Definition 5.1(iii), the first two terms are arbitrarily small for sufficiently large K and ϵ can be arbitrarily small. Therefore, equation (S.8) follows. □

Proof of Lemma S.6. Just notice that the set of function $\mathcal{H}=\{h_{a,\eta}-h_{b,\eta}: a\le b\in\mathbb{R}\}$ is a collection of bounded Lipschitz functions with Lipschitz constant $\gamma\le 2/\eta$. Then by Lemma S.5,

$$\begin{aligned}
&\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\sup_{a\le b\in\mathbb{R}}\big[\mathbb{P}_F\big(g(\mathbb{G}_{n,F})\in[a,b]\big)-\mathbb{P}\big(g(\mathbb{G}_F)\in[a-\eta,b+\eta]\big)\big]\\
&\quad\le\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\sup_{a\le b\in\mathbb{R}}\big[\mathbb{E}_F(h_{a-\eta,\eta}-h_{b,\eta})\circ g(\mathbb{G}_{n,F})-\mathbb{P}\big(g(\mathbb{G}_F)\in[a-\eta,b+\eta]\big)\big]\\
&\quad\to\sup_{F\in\mathcal{F}}\sup_{g\in\mathcal{G}}\sup_{a\le b\in\mathbb{R}}\big[\mathbb{E}_F(h_{a-\eta,\eta}-h_{b,\eta})\circ g(\mathbb{G}_F)-\mathbb{P}\big(g(\mathbb{G}_F)\in[a-\eta,b+\eta]\big)\big]\le 0.
\end{aligned}$$

□

Proof of Lemma S.7. Without loss of generality, assume $\Gamma_{\kappa,F,m,R}(\omega_1)>\Gamma_{\kappa,F,m,R}(\omega_2)$. By the definition of $\Gamma_{\kappa,F,m,R}$, $\forall\epsilon>0$, $\exists\tilde{\theta}\in\Theta_{m,F}(R)$ and $\exists\theta\in R^{\kappa}_{m,F}(\tilde{\theta})$ such that

$$\int_{\mathbf{T}}\left|\omega_2(\tilde{\theta},\mathbf{t})+\sqrt{m}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\right|^2d\mu(\mathbf{t})<\Gamma_{\kappa,R,F,m}(\omega_2)+\epsilon. \tag{S.17}$$

Therefore,

$$\begin{aligned}
&\Gamma_{\kappa,F,m,R}(\omega_1)-\Gamma_{\kappa,F,m,R}(\omega_2)\\
&\quad\le\int_{\mathbf{T}}\left|\omega_1(\tilde{\theta},\mathbf{t})+\sqrt{m}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\right|^2d\mu(\mathbf{t})\\
&\qquad-\int_{\mathbf{T}}\left|\omega_2(\tilde{\theta},\mathbf{t})+\sqrt{m}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\right|^2d\mu(\mathbf{t})+\epsilon\\
&\quad\le d_\infty(\omega_1,\omega_2)\int_{\mathbf{T}}\left|\omega_1(\tilde{\theta},\mathbf{t})+2\sqrt{m}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]+\omega_2(\tilde{\theta},\mathbf{t})\right|d\mu(\mathbf{t})+\epsilon\\
&\quad\le 2d_\infty(\omega_1,\omega_2)\int_{\mathbf{T}}\left|\omega_2(\tilde{\theta},\mathbf{t})+\sqrt{m}\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\tilde{\theta})}{d\theta}[\theta-\tilde{\theta}]\right|d\mu(\mathbf{t})\\
&\qquad+d_\infty(\omega_1,\omega_2)\int_{\mathbf{T}}\left|\omega_1(\tilde{\theta},\mathbf{t})-\omega_2(\tilde{\theta},\mathbf{t})\right|d\mu(\mathbf{t})+\epsilon.
\end{aligned}$$

Because $\tilde{\theta} \in R^{\kappa}_{m,F}(\tilde{\theta})$, $\Gamma_{\kappa,R,F,m}(\omega_2) \le \int_{\mathbf{T}} |\omega_2(\tilde{\theta}, \mathbf{t})|^2 \, d\mu(\mathbf{t}) \le \|\omega_2\|^2_\infty$. Therefore, by equation (S.17) and Jensen's inequality,

$$\begin{aligned}
&\int_{\mathbf{T}} \left| \omega_2(\tilde{\theta}, \mathbf{t}) + \sqrt{m} \frac{d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \tilde{\theta})}{d\theta} [\theta - \tilde{\theta}] \right| d\mu(\mathbf{t}) \\
&\le \sqrt{\int_{\mathbf{T}} \left| \omega_2(\tilde{\theta}, \mathbf{t}) + \sqrt{m} \frac{d\mathbb{E}_F \rho_{\mathbf{t}}(\tilde{\theta}, W_i)}{d\theta} [\theta - \tilde{\theta}] \right|^2 d\mu(\mathbf{t})} \\
&\le \sqrt{\Gamma_{\kappa,R,F,m}(\omega_2)} + \sqrt{\epsilon} \le \|\omega_2\|_\infty + \sqrt{\epsilon} \le \max\big(\|\omega_1\|_\infty, \|\omega_2\|_\infty\big) + \sqrt{\epsilon}.
\end{aligned}$$

In addition, $\int_{\mathbf{T}} |\omega_1(\tilde{\theta}, \mathbf{t}) - \omega_2(\tilde{\theta}, \mathbf{t})| \, d\mu(\mathbf{t}) \le 2 \max(\|\omega_1\|_\infty, \|\omega_2\|_\infty)$. Therefore,

$$\Gamma_{\kappa,F,m,R}(\omega_1) - \Gamma_{\kappa,F,m,R}(\omega_2) \le 4 d_\infty(\omega_1, \omega_2)\big[\max\big(\|\omega_1\|_\infty, \|\omega_2\|_\infty\big) + \sqrt{\epsilon}\big] + \epsilon.$$

Because ϵ can be any positive number, the lemma follows. □

S.8. Supplemental lemmas

Lemma S.8. *Let f be a nonnegative function defined on Θ and d is a metric such that for some $C > 0$, $d(\theta_1, \theta_2) \le C d_s(\theta_1, \theta_2)$ for every $\theta_1, \theta_2 \in \Theta$. If $\bar{\Theta}_n = \arg\inf_{\theta \in \Theta_n \cap R} f(\theta)$, then for any $A \subset \Theta_n \cap R$ and $\epsilon > 0$, $\vec{d}(\bar{\Theta}_n, A) = \sup_{\theta \in \bar{\Theta}_n} \inf_{\tilde{\theta} \in A} d(\theta, \tilde{\theta}) > \epsilon$ only if*

$$\inf_{\theta \in \Theta_n \cap R, d(\theta, A) > \epsilon} f(\theta) \le \inf_{\theta \in \Theta_n \cap R} f(\theta).$$

Proof. Because $\vec{d}(\bar{\Theta}_n, A) > \epsilon$, there exists $\bar{\theta} \in \bar{\Theta}_n$ such that $d(\bar{\theta}, A) > \epsilon + \delta$ for some $\delta > 0$. By the definition of arg inf, there exists a sequence $\{\theta_k\}_{k=1}^\infty$ in $\Theta_n \cap R$ such that $d_s(\bar{\theta}, \theta_k) \to 0$ and $\lim_{k \to \infty} f(\theta_k) = \inf_{\theta \in \Theta_n \cap R} f(\theta)$. Therefore, for any sufficiently large k,

$$d(\theta_k, A) > d(\bar{\theta}, A) - d(\bar{\theta}, \theta_k) > \epsilon + \delta - C d_s(\bar{\theta}, \theta_k) > \epsilon.$$

As a result,

$$\inf_{\theta \in \Theta_n \cap R, d(\theta, A) > \epsilon} f(\theta) \le \lim_{k \to \infty} f(\theta_k) = \inf_{\theta \in \Theta_n \cap R} f(\theta).$$ □

Lemma S.9. *There exist random variables $A_{n,F}$ and $B_{n,F}$ such that uniformly in $F \in \mathcal{F}$, $A_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})$, $B_{n,F} = O_{\mathbb{P}_F}(n^{-1})$, and*

$$\big|Q_F(\theta) - Q_n(\theta)\big| \le 2 d_{w,F}(\theta, \Theta_F) A_{n,F} + B_{n,F} = O_{\mathbb{P}_F}\big(n^{-1/2}\big).$$

Proof. First notice that

$$\begin{aligned}
\big|Q_F(\theta) - Q_n(\theta)\big| &= \left| \frac{2 \int_{\mathbf{T}} \mathbb{G}_{n,F}(\theta, \mathbf{t}) \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) \, d\mu(\mathbf{t})}{\sqrt{n}} + \frac{\int_{\mathbf{T}} \mathbb{G}_{n,F}(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t})}{n} \right| \\
&\le \frac{2 d_{w,F}(\theta, \Theta_F) \sqrt{\int_{\mathbf{T}} \mathbb{G}_{n,F}(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t})}}{\sqrt{n}} + \frac{\int_{\mathbf{T}} \mathbb{G}_{n,F}(\theta, \mathbf{t})^2 \, d\mu(\mathbf{t})}{n}
\end{aligned}$$

$$\leq 2d_{w,F}(\theta,\Theta_F)\frac{\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}\left|\mathbb{G}_{n,F}(\theta,\mathbf{t})\right|}{\sqrt{n}}+\frac{\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}\left|\mathbb{G}_{n,F}(\theta,\mathbf{t})\right|^2}{n}$$

$$= 2d_{w,F}(\theta,\Theta_F)A_{n,F}+B_{n,F},$$

where $A_{n,F} = \sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\mathbb{G}_{n,F}(\theta,\mathbf{t})|/\sqrt{n}$ and $B_{n,F} = \sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\mathbb{G}_{n,F}(\theta,\mathbf{t})|^2/n$. By Definition 5.1(iii), $\sup_{(\theta,\mathbf{t})\in\Theta\times\mathbf{T}}|\mathbb{G}_{n,F}(\theta,\mathbf{t})| = O_{\mathbb{P}_F}(1)$ uniformly in $F\in\mathcal{F}$. Consequently, $A_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})$ and $B_{n,F} = O_{\mathbb{P}_F}(n^{-1})$ uniformly in $F\in\mathcal{F}$. Lastly, notice that $d_{w,F}(\theta,\Theta_F)^2 \leq \sup_{F\in\mathcal{F}}\mathbb{E}_F\mathbf{F}(W_i)^2 < \infty$. Therefore, $2d_{w,F}(\theta,\Theta_F)A_{n,F}+B_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})$ uniformly in $F\in\mathcal{F}$. □

Lemma S.10. *$\{X_{n,a}\}_{n=1}^{\infty}$ and $\{Y_{n,a}\}_{n=1}^{\infty}$ are sequences of random variables indexed by $a\in\mathcal{A}$ and defined on the same probability space. For every $a\in\mathcal{A}$, $\mathbb{P}_a$ is a probability measure on this probability space. g is a continuous function on the real line. If $\limsup_{n\to\infty}\sup_{a\in\mathcal{A}}\mathbb{P}_a(|X_{n,a}-Y_{n,a}|>\epsilon)=0$ $\forall\epsilon>0$ and $X_{n,a}$ is asymptotically tight uniformly in $a\in\mathcal{A}$, then $\forall\epsilon>0$, $\limsup_{n\to\infty}\sup_{a\in\mathcal{A}}\mathbb{P}_a(|g(X_{n,a})-g(Y_{n,a})|>\epsilon)=0$.*

Proof. By Assumption, $\forall\eta>0$, $\exists M>0$ such that $\limsup_{n\to\infty}\sup_{a\in\mathcal{A}}\mathbb{P}_a(|X_{n,a}|>M)<\eta$. Because g is continuous, it is absolutely continuous on $[-2M,2M]$. For any $\epsilon>0$, there exists $0<\delta<M$ such that $|g(x)-g(y)|\leq\epsilon$ if $x,y\in[-2M,2M]$ and $|x-y|\leq\delta$,

$$\begin{aligned}\mathbb{P}_a\big(\left|g(X_{n,a})-g(Y_{n,a})\right|>\epsilon\big) &\leq \mathbb{P}_a\big(|X_{n,a}-Y_{n,a}|>\delta,|X_{n,a}|\leq M,|Y_{n,a}|\leq 2M\big)\\ &\qquad+\mathbb{P}_a\big(|X_{n,a}|\leq M,|Y_{n,a}|>2M\big)+\mathbb{P}_a\big(|X_{n,a}|>M\big)\\ &\leq \mathbb{P}_a\big(|X_{n,a}-Y_{n,a}|>\delta\big)\\ &\qquad+\mathbb{P}_a\big(|X_{n,a}-Y_{n,a}|>M\big)+\mathbb{P}_a\big(|X_{n,a}|>M\big).\end{aligned}$$

The first two terms converge to 0 uniformly in $a\in\mathcal{A}$. Therefore,

$$\limsup_{n\to\infty}\sup_{a\in\mathcal{A}}\mathbb{P}_a\big(\left|g(X_{n,a})-g(Y_{n,a})\right|>\epsilon\big)\leq\limsup_{n\to\infty}\sup_{a\in\mathcal{A}}\mathbb{P}_a\big(|X_{n,a}|>M\big)\leq\eta.$$

This concludes the proof because η can be any positive number. □

References

Fan, Y., M. He, and T. Li (2015), "A sensitivity analysis in the affiliated private value auction model with incomplete set of bids." Unpublished manuscript, University of Washington. Available at https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CESG2015&paper_id=59. [1]

Gandhi, A., Z. Lu, and X. Shi (2013), "Estimating demand for differentiated products with error in market shares: A moment inequalities approach." Unpublished manuscript, University of Wisconsin-Madison. Available at https://www.ssc.wisc.edu/~agandhi/homepage/Amit_Gandhi_files/gandhi.lu.shi_01-31-13.pdf. [1]

Hong, S. (2017), "Inference in semiparametric conditional moment models with partial identification." *Journal of Econometrics*, 196 (1), 156–179. [14]

Komarova, T. (2013), "Partial identification in asymmetric auctions in the absence of independence." *The Econometrics Journal*, 16 (1), S60–S92. [1]

Linton, O., K. Song, and Y.-J. Whang (2010), "An improved bootstrap test of stochastic dominance." *Journal of Econometrics*, 154 (2), 186–202. [17, 19, 20]

Santos, A. (2012), "Inference in nonparametric instrumental variables with partial identification." *Econometrica*, 80 (1), 213–275. [3]

Van Der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York. [4, 5, 6, 11]

Zhu, Y. and S. Grundl (2014), "Nonparametric tests in moment equality models with an application to infer risk aversion in first-price auctions." Unpublished manuscript. Available at https://drive.google.com/file/d/0B5Rw7pzGrkeXWnFYalFEcnFMUU0/view. [1]

Co-editor Andres Santos handled this manuscript.

Manuscript received 5 December, 2017; final version accepted 29 October, 2019; available online 20 November, 2019.