Online Appendix:
Confidence Intervals for
Projections of Partially Identified Parameters

Hiroaki Kaido* Francesca Molinari† Jörg Stoye‡

October 25, 2017

Contents

*Department of Economics, Boston University, hkaido@bu.edu. Financial support from NSF grant SES-1357643 is gratefully acknowledged.

†Department of Economics, Cornell University, fm72@cornell.edu. Financial support from NSF grant SES-0922330 is gratefully acknowledged.

‡Department of Economics, Cornell University, stoye@cornell.edu. Financial support from NSF grant SES-1260980 is gratefully acknowledged.

Structure of the Appendix

Section B states and proofs Theorem B.1, which establishes convergence-related results for our E-A-M algorithm. It also provides provides background material for the E-A-M algorithm, and details on the root-finding algorithm that we use to compute $\hat{c}_n(\theta)$. Section C verifies some of our main assumptions for moment (in)equality models that have received much attention in the literature. Section D summarizes the notation we use and the structure of the proof of Theorem 4.1,[1] and provides a proof of Theorems 4.1 (both under our main assumptions and under a high level assumption replacing Assumption 4.3 and dropping the ρ-box constraints), 4.2, 4.3 and 4.4. Section E contains the statements and proofs of the lemmas used to establish Theorems 4.1 and B.1, as well as a rigorous derivation of the almost sure representation result for the bootstrap empirical process that we use in the proof of Theorem 4.1. Section F provides further results comparing our calibrated projection method and the profiling method proposed by Bugni, Canay, and Shi (2017, BCS-profiling henceforth), and gives an example of methods' failure (including calibrated projection, BCS-profiling and the method in Pakes, Porter, Ho, and Ishii (2011)) when some key assumptions are violated. Section G provides a formal comparison of our calibrated projection method and projection of the confidence set of Andrews and Soares (2010, AS henceforth).

Throughout the Appendix we use the convention $\infty \cdot 0 = 0$.

[1] Section D.1 provides in Table D.0 a summary of the notation used throughout, and in Figure D.1 and Table D.1 a flow diagram and heuristic explanation of how each lemma contributes to the proof of Theorem 4.1.

Appendix B Additional Convergence Results and Background Materials for the E-A-M algorithm and for Computation of $\hat{c}_n(\theta)$

B.1 Theorem B.1: An Approximating Critical Level Sequence for the E-A-M Algorithm

B.1.1 Assumption B.1: A Low Level Condition Yielding a Stochastic Lipschitz-Type Property for $\hat{c}_n$

In order to establish convergence of our E-A-M algorithm, we need $\hat{c}_n$ to uniformly stochastically exhibit a Lipschitz-type property so that its mollified counterpart (see equation (B.1)) is sufficiently smooth and yields valid inference. Below we provide a low level condition under which we are able to establish the Lipschitz-type property. In Appendix C.1 we verify the condition for the canonical examples in the moment (in)equality literature.

Assumption B.1: *The model $\mathcal{P}$ for P satisfies:*

(i) $|\sigma_{P,j}(\theta)^{-1}m_j(x,\theta) - \sigma_{P,j}(\theta')^{-1}m_j(x,\theta')| \leq \bar{M}(x)\|\theta-\theta'\|$ *with* $E_P[\bar{M}(X)^2] < M$ *for all* $\theta,\theta' \in \Theta$, $x \in \mathcal{X}$, $j = 1,\cdots,J$, *and there exists a function F such that* $|\sigma_{P,j}(\theta)^{-1}m_j(\cdot,\theta)| \leq F(\cdot)$ *for all* $\theta \in \Theta$ *and* $E_P[|F(X)\bar{M}(X)|^2] < M$.

(ii) φ_j *is Lipschitz continuous in* $x \in \mathbb{R}$ *for all* $j = 1,\ldots,J$.

B.1.2 Statement and Proof of Theorem B.1

For all $\tau > 0$ let $\hat{c}_{n,\tau}(\theta)$ be a mollified version of $\hat{c}_n(\theta)$, i.e.:

$$\hat{c}_{n,\tau}(\theta) = \int_{\mathbb{R}^d} \hat{c}_n(\theta-\nu)\phi_\tau(\nu)d\nu = \int_{\mathbb{R}^d} \hat{c}_n(\theta)\phi_\tau(\theta-\nu)d\nu, \tag{B.1}$$

where the family of functions ϕ_τ is a mollifier as defined in Rockafellar and Wets (2005, Example 7.19). Choose it to be a family of bounded, measurable, smooth functions such that $\phi_\tau(z) \geq 0\ \forall z \in \mathbb{R}^d$, $\int_{\mathbb{R}^d}\phi_\tau(z)dz = 1$ and with $\mathbb{B}_\tau = \{z : \phi_\tau(z) > 0\} = \{z : \|z\| \leq \tau\}$.

Theorem B.1: *Suppose Assumptions 4.1, 4.2, 4.4, 4.5 and B.1 hold. Let τ_n be a positive sequence such that $\tau_n = n^{-\zeta}$ with $\zeta > 1/2$. Let $\{\beta_n\}$ be a positive sequence such that $\beta_n = o(1)$ and $\|\hat{D}_n - D_P\|_\infty = O_{\mathcal{P}}(\beta_n)$. Let $\varepsilon_n = \kappa_n^{-1}\sqrt{n}\tau_n \vee \beta_n$. Then,*

1.

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(\sup_{\|\theta-\theta'\|\leq\tau_n}|\hat{c}_n(\theta)-\hat{c}_n(\theta')| > C\varepsilon_n\right) = 0; \tag{B.2}$$

2. Let $\hat{c}_{n,\tau_n}$ be defined as in (B.1) with τ_n replacing τ. Then there exists $C > 0$ such that

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}} P\Big(\|\hat{c}_n - \hat{c}_{n,\tau_n}\|_\infty \leq C\varepsilon_n\Big) = 1; \tag{B.3}$$

3. There exists $R > 0$ such that $\|\hat{c}_{n,\tau_n}\|_{\mathcal{H}_\beta} \leq R$.

4. Let Assumption 4.3 also hold. Let $\{P_n, \theta_n\}$ be a sequence such that $P_n \in \mathcal{P}$ and $\theta_n \in \Theta_I(P_n)$ for all n and $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}$, $j = 1, \ldots, J$, $\Omega_{P_n} \stackrel{u}{\to} \Omega$, and $D_{P_n}(\theta_n) \to D$. Let

$$\hat{c}_{n,\rho,\tau}(\theta) \equiv \inf_{\lambda \in B^d_{n,\rho}} \hat{c}_{n,\tau}(\theta + \frac{\lambda\rho}{\sqrt{n}}). \tag{B.4}$$

For $c \geq 0$, let $U_n(\theta_n, c)$ be defined as in (D.25). Then,

$$\liminf_{n\to\infty} P_n\left(U_n(\theta_n, \hat{c}_{n,\rho,\tau_n}) \neq \emptyset\right) \geq 1 - \alpha. \tag{B.5}$$

Proof. We establish each part of the theorem separately.

Part 1. Throughout, let $C > 0$ denote a positive constant, which may be different in different appearances. Define the event

$$E_n \equiv \big\{x^\infty \in \mathcal{X}^\infty : \|\hat{D}_n - D_P\|_\infty \leq C\beta_n, \sup_{\|\theta-\theta'\|\leq\tau_n} \|\mathbb{G}_n(\theta) - \mathbb{G}_n(\theta'))\| \leq (\ln n)^2\tau_n,$$
$$\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| \leq C/\sqrt{n}, \max_{j=1,\cdots,J} \sup_{\|\theta-\theta'\|<\tau_n} |\eta_{n,j}(\theta) - \eta_{n,j}(\theta')| \leq C\tau_n\big\}. \tag{B.6}$$

Note that $(\ln n)^2\tau_n/(-\tau_n \ln \tau_n) = (\ln n)^2/\zeta \ln n = \ln n/\zeta$, and hence tends to ∞. By Assumption B.1-(i) and arguing as in the proof of Theorem 2 in Andrews (1994), condition (E.216) in Lemma E.11 is satisfied with $v = d$. Also, by Lemma E.13, (E.217) in Lemma E.11 holds with $\gamma = 1$. This therefore ensures the conditions of Lemma E.11.

Similarly, by Assumption B.1-(i) $m_j^2(x,\theta)/\sigma^2_{P,j}(\theta)$ satisfies

$$\Big|\frac{m_j^2(x,\theta)}{\sigma^2_{P,j}(\theta)} - \frac{m_j^2(x,\theta)}{\sigma^2_{P,j}(\theta)}\Big| \leq \Big|\frac{m_j(x,\theta)}{\sigma_{P,j}(\theta)} + \frac{m_j(x,\theta')}{\sigma_{P,j}(\theta')}\Big|\Big|\frac{m_j(x,\theta)}{\sigma_{P,j}(\theta)} - \frac{m_j(x,\theta')}{\sigma_{P,j}(\theta')}\Big| \tag{B.7}$$
$$\leq 2F(x)\bar{M}(x)\|\theta - \theta'\|. \tag{B.8}$$

Let $\bar{F}(x) \equiv 2F(x)\bar{M}(x)$. By Theorem 2.7.11 in van der Vaart and Wellner (2000),

$$N_{[]}(\epsilon\|\bar{F}\|_{L^2_P}, \mathcal{M}^2_P, \|\cdot\|_{L^2_P}) \leq N(\epsilon, \Theta, \|\cdot\|) \leq (\mathrm{diam}(\Theta)/\epsilon)^d, \tag{B.9}$$

where $N(\epsilon, \Theta, \|\cdot\|)$ is the covering number of Θ. This ensures

$$\int_0^\infty \sup_{P\in\mathcal{P}} \sqrt{\ln N_{[]}(\epsilon\|\bar{F}\|_{L^2_P}, \mathcal{M}^2_P, \|\cdot\|_{L^2_P})}d\epsilon < \infty. \tag{B.10}$$

Further, for any $C > 0$

$$E_P[\bar{F}^2(X)1\{\bar{F}(X) > C\}] \leq E_P[\bar{F}^2(X)]P(\bar{F}(X) > C)$$
$$\leq 4E_P[|F(X)M(X)|^2]\frac{\|\bar{F}\|_{L^1_P}}{C} \leq \frac{4M^2}{C}, \tag{B.11}$$

which implies $\lim_{C\to\infty}\sup_{P\in\mathcal{P}} E_P[\bar{F}^2(X)1\{\bar{F}(X) > C\}] = 0$. By Theorems 2.8.4 and 2.8.2 in van der Vaart and Wellner (2000), this implies that $\mathcal{S}_P$ is Donsker and pre-Gaussian uniformly in $P \in \mathcal{P}$. This therefore ensures the conditions of Lemma E.12 (i). Note also that Assumption B.1-(i) ensures the conditions of Lemma E.12 (ii). Therefore, by Lemmas E.11-E.12 and Assumption 4.4, for any $\eta > 0$, there exists $C > 0$ such that $\inf_{P\in\mathcal{P}} P(E_n) \geq 1 - \eta$ for all n sufficiently large.

Let $\theta, \theta' \in \Theta$. For each j, we have

$$\Big|\mathbb{G}^b_{n,j}(\theta) + \rho\hat{D}_{n,j}(\theta)\lambda + \varphi_j(\hat{\xi}_{n,j}(\theta)) - \mathbb{G}^b_{n,j}(\theta') - \rho\hat{D}_{n,j}(\theta')\lambda - \varphi_j(\hat{\xi}_{n,j}(\theta'))\Big|$$
$$\leq |\mathbb{G}^b_{n,j}(\theta) - \mathbb{G}^b_{n,j}(\theta')| + \rho\|\hat{D}_{n,j}(\theta) - \hat{D}_{n,j}(\theta')\| \sup_{\lambda\in B^d}\|\lambda\| + |\varphi_j(\hat{\xi}_{n,j}(\theta)) - \varphi_j(\hat{\xi}_{n,j}(\theta'))|. \quad \text{(B.12)}$$

Assume that the sample path $\{X_i\}_{i=1}^\infty$ is such that the event E_n holds. Conditional on $\{X_i\}_{i=1}^\infty$ and using $\mathbb{G}^b_{n,j}(\theta) - \mathfrak{G}^b_{n,j}(\theta) = \mathfrak{G}^b_{n,j}(\theta)\eta_{n,j}(\theta)$,

$$|\mathbb{G}^b_{n,j}(\theta) - \mathbb{G}^b_{n,j}(\theta')| \leq |\mathfrak{G}^b_{n,j}(\theta) - \mathfrak{G}^b_{n,j}(\theta')| + 2\sup_{\theta\in\Theta}|\mathfrak{G}^b_{n,j}(\theta)|\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)|$$
$$\leq |\mathfrak{G}^b_{n,j}(\theta) - \mathfrak{G}^b_{n,j}(\theta')| + 2\sup_{\theta\in\Theta}|\mathfrak{G}^b_{n,j}(\theta)|\frac{C}{\sqrt{n}}. \quad \text{(B.13)}$$

Define the event $F_n \in \mathcal{C}$ for the bootstrap weights by

$$F_n \equiv \big\{m_n \in Q : \sup_{\|\theta-\theta'\|\leq\tau_n}\|\mathfrak{G}^b_n(\theta) - \mathfrak{G}^b_n(\theta')\| \leq (\ln n)^2\tau_n,\ \sup_{\theta\in\Theta}\|\mathfrak{G}^b_n(\theta)\| \leq C\big\}. \quad \text{(B.14)}$$

By Lemma E.11 (ii) and the asymptotic tightness of $\mathfrak{G}^b_n$, for any $\eta > 0$, there exists a C such that $P^*_n(F_n) \geq 1-\eta$ for all n sufficiently large. Suppose that the multinomial bootstrap weight M_n is such that F_n holds. Then, the right hand side of (B.13) is bounded by $(\ln n)^2\tau_n + C/\sqrt{n}$ for some $C > 0$.

Next, by the triangle inequality and Assumption 4.4,

$$\|\hat{D}_{n,j}(\theta) - \hat{D}_{n,j}(\theta')\| \leq \|\hat{D}_{n,j}(\theta) - D_{P,j}(\theta)\| + \|D_{P,j}(\theta) - D_{P,j}(\theta')\| + \|\hat{D}_{n,j}(\theta') - D_{P,j}(\theta')\|$$
$$\leq C\beta_n + C\tau_n. \quad \text{(B.15)}$$

Finally, note that by the Lipschitzness of φ_j, $|\varphi_j(\hat{\xi}_{n,j}(\theta)) - \varphi_j(\hat{\xi}_{n,j}(\theta'))| \leq C|\hat{\xi}_{n,j}(\theta) - \hat{\xi}_{n,j}(\theta')|$ and

$$\hat{\xi}_{n,j}(\theta) - \hat{\xi}_{n,j}(\theta')$$
$$= \kappa_n^{-1}\Big[\sqrt{n}\Big(\frac{\bar{m}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}(1+\eta_{n,j}(\theta)) - \frac{E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big) - \sqrt{n}\Big(\frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}(1+\eta_{n,j}(\theta')) - \frac{E_P[m_j(X,\theta')]}{\sigma_{P,j}(\theta')}\Big)\Big]$$
$$+ \kappa_n^{-1}\sqrt{n}\Big(\frac{E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)} - \frac{E_P[m_j(X,\theta')]}{\sigma_{P,j}(\theta')}\Big). \quad \text{(B.16)}$$

Hence,

$$|\hat{\xi}_{n,j}(\theta) - \hat{\xi}_{n,j}(\theta')| \leq \kappa_n^{-1}|\mathbb{G}_{n,j}(\theta) - \mathbb{G}_{n,j}(\theta')|$$
$$+ \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}\eta_{n,j}(\theta) - \frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\eta_{n,j}(\theta')\Big| + \kappa_n^{-1}\sqrt{n}D_{P,j}(\bar{\theta})\|\theta-\theta'\|. \quad \text{(B.17)}$$

By Lemma E.11, the right hand side of (B.17) can be further bounded by

$$\kappa_n^{-1}(\ln n)^2\tau_n + \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - \frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\Big||\eta_{n,j}(\theta)|$$
$$+ \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\Big||\eta_{n,j}(\theta) - \eta_{n,j}(\theta')| + C\kappa_n^{-1}\sqrt{n}\tau_n$$
$$\leq \kappa_n^{-1}(\ln n)^2\tau_n + \kappa_n^{-1}\sqrt{n}\tau_n\frac{C}{\sqrt{n}} + C\kappa_n^{-1}\sqrt{n}\tau_n + C\kappa_n^{-1}\sqrt{n}\tau_n, \quad \text{(B.18)}$$

where the last inequality follows from Condition (i) and Lemma E.12 (ii).

Combining (B.12), (B.13), (B.15), and (B.16)-(B.18), we obtain

$$\left|\mathbb{G}^b_{n,j}(\theta)+\hat{D}_{n,j}(\theta)\lambda+\varphi_j(\hat{\xi}_{n,j}(\theta))-\mathbb{G}^b_{n,j}(\theta')-\hat{D}_{n,j}(\theta')\lambda-\varphi_j(\hat{\xi}_{n,j}(\theta'))\right|\le C\varepsilon_n. \tag{B.19}$$

In particular, if $\mathbf{1}\left(\Lambda^b_n(\theta,\rho,\hat{c}_n(\theta))\cap\{p'\lambda=0\}\neq\emptyset\right)=1$, it also holds that $\mathbf{1}\big(\Lambda^b_n(\theta',\rho,\hat{c}_n(\theta)+C\varepsilon_n)\cap\{p'\lambda=0\}\neq\emptyset\big)=1$ because

$$\mathbb{G}^b_{n,j}(\theta')+\hat{D}_{n,j}(\theta')\lambda+\varphi_j(\hat{\xi}_{n,j}(\theta'))\le\mathbb{G}^b_{n,j}(\theta)+\hat{D}_{n,j}(\theta)\lambda+\varphi_j(\hat{\xi}_{n,j}(\theta))+C\varepsilon_n\le\hat{c}_n(\theta)+C\varepsilon_n,$$

Recalling that $P^*_n(F_n)\ge 1-\eta$ for all n sufficiently large, we then have

$$\begin{aligned}P^*_n&\left(\left\{\Lambda^b_n(\theta',\rho,\hat{c}_n(\theta)+C\varepsilon_n)\cap\{p'\lambda=0\}\neq\emptyset\right\}\right)\\&\ge P^*_n\left(\left\{\Lambda^b_n(\theta',\rho,\hat{c}_n(\theta)+C\varepsilon_n)\cap\{p'\lambda=0\}\neq\emptyset\right\}\cap F_n\right)\\&\ge P^*_n\left(\left\{\Lambda^b_n(\theta,\rho,\hat{c}_n(\theta))\cap\{p'\lambda=0\}\neq\emptyset\right\}\cap F_n\right)\ge 1-\alpha-\eta.\end{aligned}\tag{B.20}$$

Since η is arbitrary, we have

$$\hat{c}_n(\theta')\le\hat{c}_n(\theta)+C\varepsilon_n.$$

Reversing the roles of θ and θ' and noting that $\sup_{P\in\mathcal{P}}P(E_n)\to 0$ yields the first claim of the lemma.

Part 2. To obtain the result in equation (B.3), we use that for any $\theta,\theta'\in\Theta$ such that $\|\theta-\theta'\|\le\tau_n$, $|\hat{c}_n(\theta)-\hat{c}_n(\theta')|\le C\varepsilon_n$ with probability approaching 1 uniformly in $P\in\mathcal{P}$ by the result in Part 1. This implies

$$\begin{aligned}|\hat{c}_n(\theta)-\hat{c}_{n,\tau_n}(\theta)|&=\left|\int_{\mathbb{R}^d}\hat{c}_n(\theta-\nu)\phi_{\tau_n}(\nu)d\nu-\hat{c}_n(\theta)\right|\le\int_{\mathbb{R}^d}|\hat{c}_n(\theta-\nu)-\hat{c}_n(\theta)|\,\phi_{\tau_n}(\nu)d\nu\\&=\int_{\mathbb{B}_{\tau_n}}|\hat{c}_n(\theta-\nu)-\hat{c}_n(\theta)|\,\phi_{\tau_n}(\nu)d\nu\le C\varepsilon_n\int_{\mathbb{B}_{\tau_n}}\phi_{\tau_n}(\nu)d\nu\le C\varepsilon_n.\end{aligned}$$

Part 3. By the construction of the mollified version of the critical value, we have $\hat{c}_{n,\tau_n}\in\mathcal{C}^\infty(\Theta)$ (Adams and Fournier, 2003, Theorem 2.29). Therefore it has derivatives of all order. Using the multi-index notation, for any $s>0$ and $|\alpha|\le s$, the partial derivative $\nabla^\alpha\hat{c}_{n,\tau_n}$ is bounded by some constant $M>0$ on the compact set Θ, and hence

$$\int_\Theta|\nabla^\alpha\hat{c}_{n,\tau_n}(\theta)|^2d\upsilon(\theta)\le M\upsilon(\Theta)<\infty,$$

where υ denote the Lebesgue measure on $\mathbb{R}^d$. This ensures $\nabla^\alpha\hat{c}_{n,\tau_n}\in L^2_\upsilon(\Theta)$ for all $|\alpha|\le s$. Hence, $\hat{c}_{n,\tau_n}$ is in the Sobolev-Hilbert space $H^s(\Theta^o)$ for any $s>0$. Note that when a Matérn kernel with $\nu<\infty$ is used and $\hat{c}_{n,\tau_n}$ is continuous, Lemma 3 in Bull (2011) implies that the RKHS-norm $\|\cdot\|_{\mathcal{H}_{\bar\beta}}$ (in $\mathcal{H}_{\bar\beta}(\Theta)$) and the Sobolev-Hilbert norm $\|\cdot\|_{H^{\nu+d/2}}$ are equivalent. Hence, there is $R>0$ such that $\|\hat{c}_{n,\tau_n}\|_{\mathcal{H}_\beta}\le C\|\hat{c}_{n,\tau_n}\|_{H^{\nu+d/2}}\le R$.

Part 4. By Part 2 and the definition of $\hat{c}_{n,\rho,\tau}$ in (B.4), it follows that

$$\begin{aligned}\hat{c}_{n,\rho,\tau_n}(\theta_n)&\ge\hat{c}_{n,\rho}(\theta_n)-e_n\\&\ge c^I_{n,\rho}(\theta_n)-e_n,\end{aligned}\tag{B.21}$$

for some $e_n=O_{\mathcal{P}}(\varepsilon_n)$, where the second inequality follows from the construction of $c^I_{n,\rho}$ in the proof of Lemma E.1. Note that Lemma E.3 and the fact that $\varepsilon_n=o_{\mathcal{P}}(1)$ by Part 1 imply $c^I_{n,\rho}(\theta_n)-e_n\xrightarrow{P_n}c^*_{\pi^*}$. Replicate equation (E.22) with $\hat{c}_{n,\rho,\tau_n}$ replacing $\hat{c}_{n,\rho}$, and mimic the argument following (E.22) in the proof of Lemma E.1. Then, the conclusion of the lemma follows. □

B.2 The kernel of the Gaussian Process and its Associated Function Space

Following Bull (2011), we consider two commonly used classes of kernels. The first one is the Gaussian kernel, which is given by

$$K_\beta(\theta-\theta') = \exp\big(-\sum_{k=1}^{d}|(\theta_k-\theta'_k)/\beta_k|^2\big),\ \ \beta_k\in[\underline{\beta}_k,\overline{\beta}_k],\ \ k=1,\cdots,d, \tag{B.22}$$

where $0<\underline{\beta}_k<\overline{\beta}_k<\infty$ for all k. The second one is the class of Matérn kernels defined by

$$K_\beta(\theta-\theta') = \frac{2^{1-\nu}}{D(\nu)}\Big(\sqrt{2}\nu\sum_{k=1}^{d}|(\theta_k-\theta'_k)/\beta_k|^2\Big)^{\nu}k_\nu\Big(\sqrt{2}\nu\sum_{k=1}^{d}|(\theta_k-\theta'_k)/\beta_k|^2\Big),\ \ \nu\in(0,\infty),\ \ \nu\notin\mathbb{N},$$

where D is the gamma function, and k_ν is the modified Bessel function of the second kind.[2] The index ν controls the smoothness of K_β. In particular, the Fourier transform $\hat{K}_\beta(\zeta)$ of the Matérn kernel is bounded from above and below by the order of $\|\zeta\|^{-2\nu-d}$ as $\|\zeta\|\to\infty$, i.e. $\hat{K}_\beta(\zeta)=\Theta(\|\zeta\|^{-2\nu-d})$. Similarly, the Fourier transform of the Gaussian kernel satisfies $\hat{K}_\beta(\zeta)=O(\|\zeta\|^{-2\nu-d})$ for any $\nu>0$. Below, we treat the Gaussian kernel as a kernel associated with $\nu=\infty$.

Each kernel is associated with a space of functions $\mathcal{H}_\beta(\mathbb{R}^d)$, called the reproducing kernel Hilbert space (RKHS). Below, we give some background on this space and refer to Steinwart and Christmann (2008); van der Vaart and van Zanten (2008) for further details. For $D\subseteq\mathbb{R}^d$, let $K: D\times D\to\mathbb{R}$ be a symmetric and positive definite function. K is said to be a reproducing kernel of a Hilbert space $\mathcal{H}(D)$ if $K(\cdot,\theta')\in\mathcal{H}(D)$ for all $\theta'\in D$, and

$$f(\theta)=\langle f, K(\cdot,\theta)\rangle_{\mathcal{H}(D)}$$

holds for all $f\in\mathcal{H}(D)$ and $\theta\in D$. The space $\mathcal{H}(D)$ is called a reproducing kernel Hilbert space (RKHS) over D if for all $\theta\in D$, the point evaluation functional $\delta_\theta:\mathcal{H}(D)\to\mathbb{R}$ defined by $\delta_\theta(f)=f(\theta)$ is continuous. When $K(\theta,\theta')=K_\beta(\theta-\theta')$ is used as the correlation functional of the Gaussian process, we denote the associated RKHS by $\mathcal{H}_\beta(D)$. Using Fourier transforms, the norm on $\mathcal{H}_\beta(D)$ can be written as

$$\|f\|_{\mathcal{H}_\beta}\equiv\inf_{g|_D=f}\int\frac{\hat{g}(\zeta)}{\hat{K}_\beta(\zeta)}d\zeta, \tag{B.23}$$

where the infimum is taken over functions $g:\mathbb{R}^d\to\mathbb{R}$ whose restrictions to D coincide with f, and we take $0/0=0$.

The RKHS has a connection to other well-known classes of functions. In particular, when D is a Lipschitz domain, i.e. the boundary of D is locally the graph of a Lipschitz function (Tartar, 2007) and the kernel is associated with $\nu\in(0,\infty)$, $\mathcal{H}_\beta(D)$ is equivalent to the Sobolev-Hilbert space $H^{\nu+d/2}(D^o)$, which is the space of functions on D^o such that

$$\|f\|^2_{H^{\nu+d/2}}\equiv\inf_{g|_{D^o}=f}\int\frac{\hat{g}(\zeta)}{(1+\|\zeta\|^2)^{\nu+d/2}}d\zeta \tag{B.24}$$

is finite, where the infimum is taken over functions $g:\mathbb{R}^d\to\mathbb{R}$ whose restrictions to D^o coincide with f. Further, if $\nu=\infty$, $\mathcal{H}_\beta(D)$ is continuously embedded in $H^s(D^o)$ for all $s>0$ (Bull, 2011, Lemma 3).

[2] The requirement $\nu\notin\mathbb{N}$ is not essential for the convergence result. However, it simplifies some of the arguments as one can exploit the 2ν-Hölder continuity of K_β at the origin without a log factor (Bull, 2011, Assumption 4).

Theorem 3.1 requires that c has a finite RKHS norm. This is to ensure that the approximation error made by the best linear predictor c_L of the Gaussian process regression is controlled uniformly (Narcowich, Ward, and Wendland, 2003). When a Matérn kernel is used, it suffices to bound the norm in the Sobolev-Hilbert space $H^{\nu+d/2}$ to bound c's RKHS norm. We do so in Theorem B.1 by introducing a mollified version of $\hat{c}_n$.

B.3 A Reformulation of the M-step as a Nonlinear Program

In (3.13), $\theta^{(L+1)}$ is defined as the maximizer of the following maximization problem

$$\max_{\theta\in\Theta}(p'\theta - p'\theta_L^*)_+\Big(1-\Phi\Big(\frac{\bar{g}(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big), \tag{B.25}$$

where $\bar{g}(\theta)=\max_{j=1,\cdots,J} g_j(\theta)$. Since Φ is strictly increasing, one may rewrite the objective function as

$$(p'\theta - p'\theta_L^*)_+\Big(1-\max_{j=1,\cdots,J}\Phi\Big(\frac{g_j(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big) = \min_{j=1,\cdots,J}(p'\theta - p'\theta_L^*)_+\Big(1-\Phi\Big(\frac{g_j(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big).$$

Hence, $\theta^{(L+1)}$ is a solution to the maximin problem:

$$\max_{\theta\in\Theta}\min_{j=1,\cdots,J}(p'\theta - p'\theta_L^*)_+\Big(1-\Phi\Big(\frac{g_j(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big),$$

which can be solved, for example, by Matlab's `fminimax` function. It can also be rewritten as a nonlinear program:

$$\begin{aligned}&\max_{(\theta,v)\in\Theta\times\mathbb{R}} v\\ &\qquad s.t.(p'\theta - p'\theta_L^*)_+\Big(1-\Phi\Big(\frac{g_j(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big)\geq v, j=1,\cdots,J,\end{aligned}$$

which can be solved by nonlinear optimization solvers, e.g. Matlab's `fmincon` or `KNITRO`. We note that the objective function and constraints together with their gradients are available in closed form.

B.4 Root-Finding Algorithm Used to Compute $\hat{c}_n(\theta)$

This section explains in detail how $\hat{c}_n(\theta)$ in equation (3.5) is computed. For a given $\theta\in\Theta$, $P^*(\Lambda_n^b(\theta,\rho,c)\cap\{p'\lambda = 0\}\neq\emptyset)$ increases in c (with $\Lambda_n^b(\theta,\rho,c)$ defined in (3.1)), and so $\hat{c}_n(\theta)$ can be quickly computed via a root-finding algorithm, such as the Brent-Dekker Method (BDM), see Brent (1971) and Dekker (1969). To do so, define $h_\alpha(c)=\frac{1}{B}\sum_{b=1}^B\psi_b(c)-(1-\alpha)$ where

$$\psi_b(c(\theta)) = \mathbf{1}(\Lambda_n^b(\theta,\rho,c)\cap\{p'\lambda=0\}\neq\emptyset).$$

Let $\bar{c}(\theta)$ be an upper bound on $\hat{c}_n(\theta)$ (for example, the asymptotic Bonferroni bound $\bar{c}(\theta)\equiv\Phi^{-1}(1-\alpha/J)$). It remains to find $\hat{c}_n(\theta)$ so that $h_\alpha(\hat{c}_n(\theta))=0$ if $h_\alpha(0)\leq 0$. It is possible that $h_\alpha(0)>0$ in which case we output $\hat{c}_n(\theta)=0$. Otherwise, we use BDM to find the unique root to $h_\alpha(c)$ on $[0,\bar{c}(\theta)]$ where, by construction, $h_\alpha(\bar{c}_n(\theta))\geq 0$. We propose the following algorithm:

Step 0 (Initialize)

(i) Set *Tol* equal to a chosen tolerance value;

(ii) Set $c_L=0$ and $c_U=\bar{c}(\theta)$ (values of c that bracket the root $\hat{c}_n(\theta)$);

(iii) Set $c_{-1}=c_L$ and $c_{-2}=[]$ to be undefined for now (proposed values of c from 1 and 2 iterations prior). Also set $c_0=c_L$ and $c_1=c_U$.

(iv) Compute $\varphi_j(\hat{\xi}_{n,j}(\theta))$ $j = 1, \cdots, J$;

(v) Compute $\hat{D}_{P,n}(\theta)$;

(vi) Compute $\mathbb{G}_{n,j}^b$ for $b = 1, \cdots, B$, $j = 1, \cdots, J$;

(vii) Compute $\psi_b(c_L)$ and $\psi_b(c_U)$ for $b = 1, \cdots, B$;

(viii) Compute $h_\alpha(c_L)$ and $h_\alpha(c_U)$.

Step 1 (Method Selection)

Use the BDM rule to select the updated value of c, say c_2. The value is updated using one of three methods: Inverse Quadratic Interpolation, Secant, or Bisection. The selection rule is based on the values of c_i, $i = -2, -1, 0, 1$ and the corresponding function values.

Step 2 (Update Value Function)

Update the value of $h_\alpha(c_2)$. We can exploit previous computation and monotonicity function $\psi_b(c_2)$ to reduce computational time:

1. If $\psi_b(c_L) = \psi_b(c_U) = 0$, then $\psi_b(c_2) = 0$;
2. If $\psi_b(c_L) = \psi_b(c_U) = 1$, then $\psi_b(c_2) = 1$.

Step 3 (Update)

(i) If $h_\alpha(c_2) \geq 0$, then set $c_U = c_2$. Otherwise set $c_L = c_2$.

(ii) Set $c_{-2} = c_{-1}$, $c_{-1} = c_0$, $c_0 = c_L$, and $c_1 = c_U$.

(iii) Update corresponding function values $h_\alpha(\cdot)$.

Step 4 (Convergence)

(i) If $h_\alpha(c_U) \leq \mathit{Tol}$ or if $|c_U - c_L| \leq \mathit{Tol}$, then output $\hat{c}_n(\theta) = c_U$ and exit. Note: $h_\alpha(c_U) \geq 0$, so this criterion ensures that we have *at least* $1 - \alpha$ coverage.

(ii) Otherwise, return to **Step 1**.

The computationally difficult part of the algorithm is computing $\psi_b(\cdot)$ in **Step 2**. This is simplified for two reasons. First, evaluation of $\psi_b(c)$ entails determining whether a constraint set comprised of $J + 2d - 2$ linear inequalities in $d - 1$ variables is feasible. This can be accomplished efficiently employing commonly used software.[3] Second, we exploit monotonicity in $\psi_b(\cdot)$, reducing the number of linear programs needed to be solved.

Appendix C Verification of Assumptions for the Canonical Moment (In)equalities Examples

In this section we verify: (i) Assumption B.1 which is the crucial condition in Theorem B.1, and (ii) Assumption 4.3-(II), for the canonical examples in the moment (in)equalities literature:

[3] Examples of high-speed solves for linear programs include CVXGEN, availiable from `http://www.cvxgen.com` and Gurobi, available from `http://www.gurobi.com`.

1. **Mean with interval data (of which missing data is a special case)**. Here we assume that W_0, W_1 are two observable random variables such that $P(W_0 \leq W_1) = 1$. The identified set is defined as

$$\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R} : E_P(W_0) - \theta \leq 0, \theta - E_P(W_1) \leq 0\}. \tag{C.1}$$

2. **Linear regression with interval outcome data and discrete regressors.** Here the modeling assumption is that $W = Z'\theta + u$, where $Z = [Z_1; \ldots; Z_d]$ is a $d \times 1$ random vector with $Z_1 = 1$. We assume that Z has k points of support denoted $z^1, \ldots, z^k \in \mathbb{R}^d$ with $\max_{r=1,\ldots,k} \|z^r\| < M < \infty$. The researcher observes $\{W_0, W_1, Z\}$ with $P(W_0 \leq W \leq W_1 | Z = z^r) = 1, r = 1, \ldots, k$. The identified set is

$$\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R}^d : E_P(W_0|Z = z^r) - z^{r\prime}\theta \leq 0, z^{r\prime}\theta - E_P(W_1|Z = z^r) \leq 0, r = 1, \ldots, k\}. \tag{C.2}$$

3. **Best linear prediction with interval outcome data and discrete regressors.** Here the variables are defined as for the linear regression case. Beresteanu and Molinari (2008) show that the identified set for the parameters of a best linear predictor of W conditional on Z is given by the set $\Theta_I(P) = E_P(ZZ')^{-1}E_P(Z\mathbf{W})$, where $\mathbf{W} = [W_0, W_1]$ is a random closed set and, with some abuse of notation, $E_P(Z\mathbf{W})$ denotes the Aumann expectation of $Z\mathbf{W}$.

   Here we go beyond the results in Beresteanu and Molinari (2008) and derive a moment inequality representation for $\Theta_I(P)$ when Z has a discrete distribution. We denote by u^r the vector $u^r = e^{r\prime}(M_P'M_P)^{-1}M_P'E_P(ZZ')$, $r = 1, \ldots, k$, where e^r is the r-th basis vector in $\mathbb{R}^k$ and M_P is a $d \times K$ matrix with r-th column equal to $P(Z = z^r)z^r$; we let $q^r = u^r E_P(ZZ')^{-1}$. Observe that for any selection $\tilde{W} \in \mathbf{W}$ $a.s.$ one has $u^r E_P(ZZ')^{-1}E_P(Z\tilde{W}) = e^{r\prime}[E_P(\tilde{W}|Z = z^1); \ldots; E_P(\tilde{W}|Z = z^k)]$, so that the support function in direction u^r is maximized/minimized by setting $E_P(\tilde{W}|Z = z^r)$ equal to $E_P(W_1|Z = z^r)$ and $E_P(W_0|Z = z^r)$, respectively. Hence, the identified set can be written in terms of moment inequalities as

$$\begin{aligned}\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R}^d : & \; q^r[E_P(Z(Z'\theta - W_0 - \mathbf{1}(q^r Z > 0)(W_1 - W_0)))] \leq 0 \\ & - q^r[E_P(Z(Z'\theta - W_0 - \mathbf{1}(q^r Z < 0)(W_1 - W_0)))] \leq 0, r = 1, \ldots, k\}.\end{aligned} \tag{C.3}$$

   The set is expressed through evaluation of its support function, given in Bontemps, Magnac, and Maurin (2012, Proposition 2), at directions $\pm u^r$; these are the directions orthogonal to the flat faces of $\Theta_I(P)$.

4. **Complete information entry games with pure strategy Nash equilibrium as solution concept.** Here again we assume that the vector Z has k points of support with bounded norm, and the identified set is

$$\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R}^d : \text{ equations (5.1), (5.2), (5.3), (5.4) hold for all } Z = z^r, r = 1, \ldots, k\}. \tag{C.4}$$

In the first three examples we let $X \equiv (W_0, W_1, Z)'$. In the last example we let $X \equiv (Y_1, Y_2, Z)'$. Throughout, we propose to estimate $E_P(W_\ell|Z = z^r)$ and $E_P(Y_1 = s, Y_2 = t|Z = z^r)$, $\ell = 0, 1$, $(s, t) \in \{0, 1\} \times \{0, 1\}$ and $r = 1, \ldots, k$, using

$$\hat{E}_n(W_\ell|Z = z^r) = \frac{\sum_{i=1}^n W_{\ell,i}\mathbf{1}(Z_i = z^r)}{\sum_{i=1}^n \mathbf{1}(Z_i = z^r)}, \tag{C.5}$$

$$\hat{E}_n(Y_1 = s, Y_2 = t|Z = z^r) = \frac{\sum_{i=1}^n \mathbf{1}(Y_{1,i} = s, Y_{2,i} = t, Z_i = z^r)}{\sum_{i=1}^n \mathbf{1}(Z_i = z^r)}, \tag{C.6}$$

as it is done in, e.g., Ciliberto and Tamer (2009). We assume that for each of the four canonical examples under consideration, Assumption 4.1 as well as one of the assumptions below hold.

Assumption C.1: *The model* $\mathcal{P}$ *for* P *satisfies* $\min_{\ell=0,1}\min_{r=1,\dots,k} Var_P(W_\ell|Z=z^r) > \underline{\sigma} > 0$ *and* $\min_{r=1,\dots,k} P(Z=z^r) > \varpi > 0$.

Assumption C.2: *The model* $\mathcal{P}$ *for* P *satisfies: (1)* $\text{eig}(M_P'M_P) > \varsigma$*; (2)* $\text{eig}(E_P(ZZ')) > \varsigma$*; (3)* $\text{eig}(Corr_P([vech(ZZ');W_0])) > \varsigma$ *and* $\text{eig}(Corr_P([vech(ZZ');W_1])) > \varsigma$*; for some* $\varsigma > 0$*, where* $vech(A)$ *denotes the half-vectorization of the matrix* A.

Assumption C.3: *The model* $\mathcal{P}$ *for* P *satisfies* $\min_{r=1,\dots,k,(s,t)\in\{0,1\}\times\{0,1\}} P(Y_1=s,Y_2=t,Z=z^r) > \varpi > 0$.

These are simple to verify low level conditions. We note that Imbens and Manski (2004) and Stoye (2009) directly assume the unconditional version of C.1, while Beresteanu and Molinari (2008) assume C.1 itself.

C.1 Verification of Assumption B.1 in Theorem B.1

We show that in each of the four examples $\frac{m_j(x,\theta)}{\sigma_{P,j}(\theta)}$, $j=1,\dots,J$ is Lipschitz continuous in $\theta\in\Theta$ for all $x\in\mathcal{X}$ and that D_P can be estimated at rate $n^{-1/2}$.

1. **Mean with interval data**. Here $\sigma_{P,\ell}(\theta)=\sigma_{P,\ell}$, and under Assumption C.1 it is uniformly bounded from below. Then

$$\left|\frac{m_j(x,\theta)}{\sigma_{P,j}} - \frac{m_j(x,\theta')}{\sigma_{P,j}}\right| = \frac{\|(\theta'-\theta)\|}{\sigma_{P,j}(\theta)}, \quad \ell=0,1,$$
$$D_{P,\ell}(\theta) = \frac{(-1)^{(1-\ell)}}{\sigma_{P,\ell}}, \quad \ell=0,1.$$

   Assumption C.1 then guarantees that Assumption B.1 is satisfied.

2. **Linear regression with interval outcome data and discrete regressors.** Here again $\sigma_{P,\ell r}(\theta)=\sigma_{P,\ell r}$, and under Assumptions C.1-C.2 it is uniformly bounded from below. We first consider the rescaled function $\frac{(-1)^j(W_\ell \mathbf{1}(Z=z^r)/P(Z=z^r)-z^{r\prime}\theta)}{\sigma_{P,\ell r}}$:

$$\left|\frac{(-1)^j(W_\ell \mathbf{1}(Z=z^r)/P(Z=z^r)-z^{r\prime}\theta)}{\sigma_{P,\ell r}} - \frac{(-1)^j(W_\ell \mathbf{1}(Z=z^r)/P(Z=z^r)-z^{r\prime}\theta')}{\sigma_{P,\ell r}}\right| = \|z^r\|\frac{\|(\theta'-\theta)\|}{\sigma_{P,\ell r}(\theta)}, \quad \ell=0,1,$$

   so that Assumption B.1 is satisfied for these rescaled functions by Assumptions C.1-C.2. Next, we observe that

$$D_{P,j} = \frac{(-1)^{(1-j)}z^{r\prime}}{\sigma_{P,\ell r}}, \quad \ell=0,1, r=1,\dots,k,$$

   and it can be estimated at rate $n^{-1/2}$ by Lemma E.12. Theorem B.1 then holds observing that $|P(Z=z^r)/\sum_{i=1}^n \mathbf{1}(Z_i=z^r)-1| = O_{\mathcal{P}}(n^{-1/2})$ and treating this random element similarly to how we treat $\eta_{n,j}(\cdot)$ in the proof of Theorem B.1.

3. **Best linear prediction with interval outcome data and discrete regressors.** Here

$$m_r(X_i,\theta) = q^r[Z_i(Z_i'\theta - (W_{0,i}+\mathbf{1}(q^rZ_i>0)(W_{1,i}-W_{0,i})))] \tag{C.7}$$

   hence is Lipschitz in θ with constant Z_iZ_i'. Under Assumptions C.1-C.2, $Var_P(m_r(X_i,\theta))$ is uniformly bounded from below, and Lipschitz in θ with a constant that depends on Z_i^4. Hence $\frac{m_r(X_i,\theta)}{\sigma_{P,r}(\theta)}$ is Lipschitz in θ

with a constant that depends on powers of Z. Because Z has bounded support, Assumption B.1 is satisfied. A simple argument yields that D_P can be estimated at rate $n^{-1/2}$.

4. **Complete information entry games with pure strategy Nash equilibrium as solution concept.** Here again $\sigma_{P,str}(\theta) = \sigma_{P,str}$, and under Assumptions 4.1 and C.3 it is uniformly bounded from below. The result then follows from a similar argument as the one used in Example 2 (Linear regression with interval outcome data and discrete regressors), observing that the rescaled function of interest is now

$$\frac{\mathbf{1}(Y_1 = s, Y_2 = t | Z = z^r)/P(Z = z^r) - g_{str}(\theta)}{\sigma_{P,str}}, \quad (s,t) \in \{0,1\} \times \{0,1\}, r = 1, \dots, k,$$

and the gradient is

$$\frac{1}{\sigma_{P,str}} \nabla_\theta g_{str}(\theta), \quad (s,t) \in \{0,1\} \times \{0,1\}, r = 1, \dots, k,$$

where $g_{str}(\theta)$ are model-implied entry probabilities, and hence taking their values in $[0,1]$. The entry models typically posited assume that payoff shocks have smooth distributions (e.g., multivariate normal), yielding that $\nabla_\theta g_{str}(\theta)$ is well defined and bounded.

C.2 Verification of Assumption 4.3-(II)

Here we verify Assumption 4.3-(II) for the canonical examples in the moment (in)equalities literature:

1. **Mean with interval data.** In the generalization of this example in Imbens and Manski (2004) and Stoye (2009), equations (4.1)-(4.2) are satisfied by construction, equation (4.3) is directly assumed.

2. **Linear regression with interval outcome data and discrete regressors.** Equation (4.1) is satisfied by construction. Given the estimator that we use for the population moment conditions, we verify equation (4.3) for the variances of the limit distribution of the vector $[\sqrt{n}(\hat{E}_n(W_\ell | Z = z^r) - E_P(W_\ell | Z = z^r))]_{\ell \in \{0,1\}, r=1,\dots,k}$. We then have that equation (4.3) follows from Assumption C.1. Concerning equation (4.3), this needs to be verified for the correlation matrix of the limit distribution of a $r \times 1$ random vector that for each $r = 1, \dots, k$ equals any choice in $\{\sqrt{n}(\hat{E}_n(W_0 | Z = z^r) - E_P(W_0 | Z = z^r)), \sqrt{n}(\hat{E}_n(W_1 | Z = z^r) - E_P(W_1 | Z = z^r))\}$, which suffices for our results to hold. We then have that (4.2) holds because the correlation matrix is diagonal.

3. **Best linear prediction with interval outcome data and discrete regressors.** Equation (4.1) is again satisfied by construction. Equation (4.2) holds under Assumptions C.1-C.2. Equation (4.3) is verified to hold under Assumption C.1 in Beresteanu and Molinari (2008, p. 808).

4. **Complete information entry games with pure strategy Nash equilibrium as solution concept.** In this case equations (5.3) and (5.4) are paired, but the corresponding moment functions differ by the model implied probability of the region of multiplicity, hence equation (4.1) is satisfied by construction. Given the estimator that we use for the population moment conditions, we verify equations (4.2) and (4.3) for the variances and for the correlation matrix of the limit distribution of the vector $\sqrt{n}(\hat{E}_n(Y_1 = s, Y_2 = t | Z = z^r) - E_P(Y_1 = s, Y_2 = t | Z = z^r)_{(s,t) \in \{0,1\} \times \{0,1\}, r=1,\dots,k})$, which suffices for our results to hold. Equation (4.2) holds provided that $|Corr(Y_{i1}(1 - Y_{i2}), Y_{i1} Y_{i2})| < 1 - \epsilon$ for some $\epsilon > 0$ and Assumption C.3 holds.[4] To see that equation (4.3) also holds, note that Assumption C.3 yields that $P(Y_{i1} = 1, Y_{i2} = 0, Z_i = z^r)$ is

[4] In more general instances with more than two players, it follows if the multinomial distribution of outcomes of the game (reduced by one element) has a correlation matrix with eigenvalues uniformly bounded away from zero.

uniformly bounded away from 0 and 1, thereby implying that for each $(s,t) \in \{0,1\} \times \{0,1\}, r = 1,\dots,k$, $(P(Y_1 = s, Y_2 = t|Z = z^r)(1 - P(Y_1 = s, Y_2 = t|Z = z^r)))/(P(Z = z^r)(1 - P(Z = z^r)))$ is uniformly bounded away from zero.

Appendix D Proof of Theorems 4.1, 4.2, 4.3 and 4.4

D.1 Notation and Structure of the Proof of Theorem 4.1

For any sequence of random variables $\{X_n\}$ and a positive sequence a_n, we write $X_n = o_{\mathcal{P}}(a_n)$ if for any $\epsilon, \eta > 0$, there is $N \in \mathbb{N}$ such that $\sup_{P \in \mathcal{P}} P(|X_n/a_n| > \epsilon) < \eta, \forall n \geq N$. We write $X_n = O_{\mathcal{P}}(a_n)$ if for any $\eta > 0$, there is a $M \in \mathbb{R}_+$ and $N \in \mathbb{N}$ such that $\sup_{P \in \mathcal{P}} P(|X_n/a_n| > M) < \eta, \forall n \geq N$.

Table D.0: Important notation. Here $(P_n, \theta_n) \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ is a subsequence as defined in (D.3)-(D.4) below, $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$, $B^d = \{x \in \mathbb{R}^d : |x_i| \leq 1, i = 1,\dots,d\}$, $B^d_{n,\rho} \equiv \frac{\sqrt{n}}{\rho}(\Theta - \theta_n) \cap B^d$, $\mathfrak{B}^d_\rho = \lim_{n\to\infty} B^d_{n,\rho}$, and $\lambda \in \mathbb{R}^d$.

$\mathbb{G}_{n,j}(\cdot)$	=	$\frac{\sqrt{n}(\bar{m}_{n,j}(\cdot) - E_P(m_j(X_i,\cdot)))}{\sigma_{P,j}(\cdot)},\ j = 1,\dots,J$	Sample empirical process.
$\mathbb{G}^b_{n,j}(\cdot)$	=	$\frac{\sqrt{n}(\bar{m}^b_{n,j}(\cdot) - \bar{m}_{n,j}(\cdot))}{\hat{\sigma}_{n,j}(\cdot)},\ j = 1,\dots,J$	Bootstrap empirical process.
$\eta_{n,j}(\cdot)$	=	$\frac{\sigma_{P,j}(\cdot)}{\hat{\sigma}_{n,j}(\cdot)} - 1,\ j = 1,\dots,J$	Estimation error in sample moments' asymptotic standard deviation.
$D_{P,j}(\cdot)$	=	$\nabla_\theta \left(\frac{E_P(m_j(X_i,\cdot))}{\sigma_{P,j}(\cdot)}\right),\ j = 1,\dots,J$	Gradient of population moments w.r.t. θ, with estimator $\hat{D}_{n,j}(\cdot)$.
$\gamma_{1,P_n,j}(\cdot)$	=	$\frac{E_{P_n}(m_j(X_i,\cdot))}{\sigma_{P_n,j}(\cdot)},\ j = 1,\dots,J$	Studentized population moments.
$\pi_{1,j}$	=	$\lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n')$	Limit of rescaled population moments, constant $\forall \theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ by Lemma E.5.
$\pi^*_{1,j}$	=	$\begin{cases} 0, & \text{if } \pi_{1,j} = 0, \\ -\infty, & \text{if } \pi_{1,j} < 0. \end{cases}$	"Oracle" GMS.
$\hat{\xi}_{n,j}(\cdot)$	=	$\begin{cases} \kappa_n^{-1}\sqrt{n}\bar{m}_{n,j}(\cdot)/\hat{\sigma}_{n,j}(\cdot), & j = 1,\dots,J_1 \\ 0, & j = J_1+1,\dots,J \end{cases}$	Rescaled studentized sample moments, set to 0 for equalities.
$\varphi^*_j(\xi)$	=	$\begin{cases} \varphi_j(\xi) & \pi_{1,j} = 0 \\ -\infty & \pi_{1,j} < 0 \\ 0 & j = J_1+1,\cdots,J. \end{cases}$	Infeasible GMS that is less conservative than φ_j.
$u_{n,j,\theta_n}(\lambda)$	=	$\{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}))$	Mean value expansion of nonlinear constraints with sample empirical process and "oracle" GMS, with $\bar{\theta}_n$ componentwise between θ_n and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$.
$U_n(\theta_n, c)$	=	$\left\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u_{n,j,\theta_n}(\lambda) \leq c,\ \forall j = 1,\dots,J\right\}$	Feasible set for nonlinear sample problem intersected with $p'\lambda = 0$.
$\mathfrak{w}_j(\lambda)$	=	$\mathbb{Z}_j + \rho D_j\lambda + \pi^*_{1,j}$	Linearized constraints with a Gaussian shift and "oracle" GMS.
$\mathfrak{W}(c)$	=	$\left\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \leq c,\ \forall j = 1,\dots,J\right\}$	Feasible set for linearized limit problem intersected with $p'\lambda = 0$.
c_{π^*}	=	$\inf\{c \in \mathbb{R}_+ : \Pr(\mathfrak{W}(c) \neq \emptyset) \geq 1 - \alpha\}$.	Limit problem critical level.
$v^b_{n,j,\theta_n'}(\lambda)$	=	$\mathbb{G}^b_{n,j}(\theta_n') + \rho\hat{D}_{n,j}(\theta_n')\lambda + \varphi_j(\hat{\xi}_{n,j}(\theta_n'))$	Linearized constraints with bootstrap empirical process and sample GMS.
$V^b_n(\theta_n', c)$	=	$\left\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap v^b_{n,j,\theta_n'}(\lambda) \leq c,\ \forall j = 1,\dots,J\right\}$	Feasible set for linearized bootstrap problem with sample GMS and $p'\lambda = 0$.
$v^I_{n,j,\theta_n'}(\lambda)$	=	$\mathbb{G}^b_{n,j}(\theta_n') + \rho\hat{D}_{n,j}(\theta_n')\lambda + \varphi^*_j(\hat{\xi}_{n,j}(\theta_n'))$	Linearized constraints with bootstrap empirical process and infeasible sample GMS.
$V^I_n(\theta_n', c)$	=	$\left\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap v^I_{n,j,\theta_n'}(\lambda) \leq c,\ \forall j = 1,\dots,J\right\}$	Feasible set for linearized bootstrap problem with infeasible sample GMS and $p'\lambda = 0$.
$\hat{c}_n(\theta)$	=	$\inf\{c \in \mathbb{R}_+ : P^*(V^b_n(\theta, c) \neq \emptyset) \geq 1 - \alpha\}$	Bootstrap critical level.
$\hat{c}_{n,\rho}(\theta)$	=	$\inf_{\lambda \in B^d_{n,\rho}} \hat{c}_n(\theta + \frac{\lambda\rho}{\sqrt{n}})$	Smallest value of the bootstrap critical level in a $B^d_{n,\rho}$ neighborhood of θ.
$\hat{\sigma}^M_{n,j}(\theta)$	=	$\hat{\mu}_{n,j}(\theta)\hat{\sigma}_{n,j}(\theta) + (1 - \hat{\mu}_{n,j}(\theta))\hat{\sigma}_{n,j+R_1}(\theta)$	Weighted sum of the estimators of the standard deviations of paired inequalities

Figure D.1: Structure of Lemmas used in the proof of Theorem 4.1.

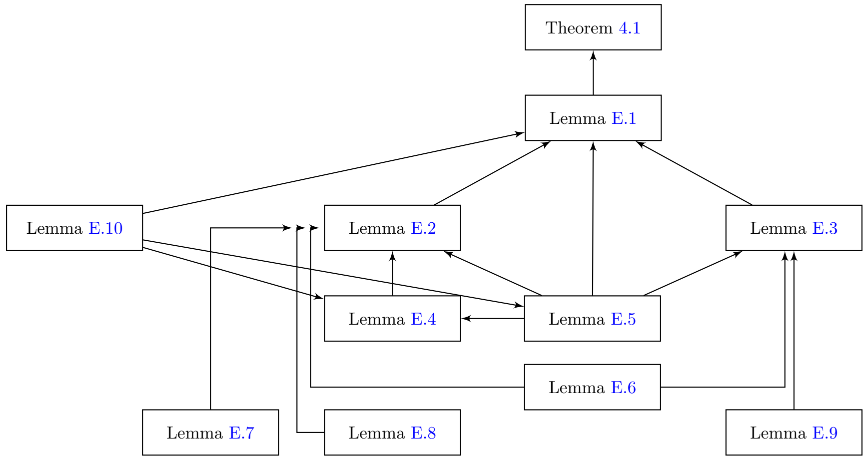

Table D.1: Heuristics for the role of each Lemma in the proof of Theorem 4.1. Notes: (i) Uniformity in Theorem 4.1 is enforced arguing along subsequences; (ii) When needed, random variables are realized on the same probability space as shown in Lemma E.1 and Lemma E.17 (see Appendix E.3 for details); (iii) Here $(P_n, \theta_n) \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ is a subsequence as defined in (D.3)-(D.4) below; (iv) All results hold for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$.

Theorem 4.1	$P_n(p'\theta_n \in CI) \geq P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}(\theta_n)) \neq \emptyset\right)$. Coverage is conservatively estimated by the probability that U_n is nonempty.
Lemma E.1	$\liminf P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}(\theta_n)) \neq \emptyset\right) \geq 1 - \alpha$.
Lemma E.2	$P_n(U(\theta_n, c_n^I(\theta_n)) \neq \emptyset, \mathfrak{W}(c_{\pi^*}) = \emptyset) + P_n(U(\theta_n, c_n^I(\theta_n)) = \emptyset, \mathfrak{W}(c_{\pi^*}) \neq \emptyset) = o_{\mathcal{P}}(1)$. Argued by comparing U_n and its limit $\mathfrak{W}$ (after coupling).
Lemma E.3	$P_n^*(V_n^I(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset) \to 0$ and $c_n^I(\theta'_n) \overset{P_n}{\to} c_{\pi^*}$ if $c_{\pi^*} > 0$. The bootstrap critical value that uses the less conservative GMS yileds a convergent critical value.
Lemma E.4	$\sup_{\lambda \in B^d} \lvert \max_j(u_{n,j,\theta_n}(\lambda) - c_n^I(\theta_n)) - \max_j(\mathfrak{w}_j(\lambda) - c_{\pi^*})\rvert = o_{\mathcal{P}}(1)$, and similarly for $\mathfrak{w}_j$ and v^I_{n,j,θ'_n}. The criterion functions entering U_n and $\mathfrak{W}$ converge to each other.
Lemma E.5	Local-to-binding constraints are selected by GMS uniformly over the ρ-box (intuition: $\rho n^{-1/2} = o_{\mathcal{P}}(\kappa_n^{-1})$), and $\lVert\hat{\xi}_n(\theta'_n) - \kappa_n^{-1}\sqrt{n}\sigma^{-1}_{P_n,j}(\theta'_n)E_{P_n}[m_j(X_i, \theta'_n)]\rVert = o_{\mathcal{P}}(1)$.
Lemma E.6	$\forall \eta > 0 \; \exists \delta > 0, : \Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}^{-\delta}(c) = \emptyset\}) < \eta$, and similarly for V_n^I. It is unlikely that these sets are nonempty but become empty upon slightly tightening stochastic constraints.
Lemma E.7	Intersections of constraints whose gradients are almost linearly dependent are unlikely to realize inside $\mathfrak{W}$. Hence, we can ignore irregularities that occur as linear dependence is approached.
Lemma E.8	If there are weakly more equality constraints than parameters, then c is uniformly bounded away from zero. This simplifies some arguments.
Lemma E.9	If two paired inequalities are local to binding, then they are also asymptotically identical up to sign. This justifies "merging" them.
Lemma E.10	$\eta_{n,j}(\cdot)$ converges to zero uniformly in P and θ.

D.2 Proof of Theorems 4.1 and 4.2

D.2.1 Main Proofs

Proof of Theorem 4.1

Following Andrews and Guggenberger (2009), we index distributions by a vector of nuisance parameters relevant for the asymptotic size. For this, let $\gamma_P \equiv (\gamma_{1,P}, \gamma_{2,P}, \gamma_{3,P})$, where $\gamma_{1,P} = (\gamma_{1,P,1}, \cdots, \gamma_{1,P,J})$ with

$$\gamma_{1,P,j}(\theta) = \sigma_{P,j}^{-1}(\theta) E_P[m_j(X_i, \theta)],\ j = 1, \cdots, J, \tag{D.1}$$

$\gamma_{2,P} = (s(p, \Theta_I(P)), vech(\Omega_P(\theta)), vec(D_P(\theta)))$, and $\gamma_{3,P} = P$. We proceed in steps.

Step 1. Let $\{P_n, \theta_n\} \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ be a sequence such that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n) = \liminf_{n\to\infty} P_n(p'\theta_n \in CI_n), \tag{D.2}$$

with $CI_n = [-s(-p, \mathcal{C}_n(\hat{c}_n)), s(p, \mathcal{C}_n(\hat{c}_n))]$. We then let $\{l_n\}$ be a subsequence of $\{n\}$ such that

$$\liminf_{n\to\infty} P_n(p'\theta_n \in CI_n) = \lim_{n\to\infty} P_{l_n}(p'\theta_{l_n} \in CI_{l_n}). \tag{D.3}$$

Then there is a further subsequence $\{a_n\}$ of $\{l_n\}$ such that

$$\lim_{a_n\to\infty} \kappa_{a_n}^{-1} \sqrt{a_n} \sigma_{P_{a_n},j}^{-1}(\theta_{a_n}) E_{P_{a_n}}[m_j(X_i, \theta_{a_n})] = \pi_{1,j} \in \mathbb{R}_{[-\infty]},\ j = 1, \ldots, J. \tag{D.4}$$

To avoid multiple subscripts, with some abuse of notation we write (P_n, θ_n) to refer to (P_{a_n}, θ_{a_n}) throughout this Appendix. We let

$$\pi_{1,j}^* = \begin{cases} 0 & \text{if } \pi_{1,j} = 0, \\ -\infty & \text{if } \pi_{1,j} < 0. \end{cases} \tag{D.5}$$

The projection of θ_n is covered when

$$\begin{aligned}
& -s(-p, \mathcal{C}_n(\hat{c}_n)) \le p'\theta_n \le s(p, \mathcal{C}_n(\hat{c}_n)) \\
\Leftrightarrow & \begin{Bmatrix} \inf p'\vartheta \\ s.t.\ \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \le \hat{c}_n(\vartheta), \forall j \end{Bmatrix} \le p'\theta_n \le \begin{Bmatrix} \sup p'\vartheta \\ s.t.\ \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \le \hat{c}_n(\vartheta), \forall j \end{Bmatrix} \\
\Leftrightarrow & \begin{Bmatrix} \inf_\lambda p'\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \le \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix} \le 0 \\
& \le \begin{Bmatrix} \sup_\lambda p'\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \le \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix} \qquad (D.6) \\
\Leftrightarrow & \begin{Bmatrix} \inf_\lambda p'\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix} \le 0 \\
& \le \begin{Bmatrix} \sup_\lambda p'\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix}, \qquad (D.7)
\end{aligned}$$

with $\eta_{n,j}(\cdot) \equiv \sigma_{P,j}(\cdot)/\hat{\sigma}_{n,j}(\cdot) - 1$ and where we localized ϑ in a $\sqrt{n}/\rho$-neighborhood of $\Theta - \theta_n$ and we took a mean value expansion yielding $\forall j$

$$\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}=\{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar{\theta}_n)\lambda+\sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})). \tag{D.8}$$

Denote $B^d_{n,\rho}\equiv\frac{\sqrt{n}}{\rho}(\Theta-\theta_n)\cap B^d$, with $B^d=\{x\in\mathbb{R}^d:|x_i|\leq 1, i=1,\dots,d\}$. The event in (D.7) is implied by

$$\Leftarrow\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ s.t.\lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar{\theta}_n)\lambda+\sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat{c}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\leq 0$$
$$\leq\left\{\begin{array}{c}\sup_\lambda p'\lambda\\ s.t.\lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar{\theta}_n)\lambda+\sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat{c}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}, \tag{D.9}$$

Step 2. This step is used only when Assumption 4.3-(II) is invoked. When this assumption is invoked, recall that in equation (2.5) we use the estimator specified in Lemma E.10 equation (E.188) for $\sigma_{P,j}, j=1,\dots,2R_1$ (with $R_1\leq J_1/2$ defined in the statement of the assumption). In equation (3.1) we use the sample analog estimators of $\sigma_{P,j}$ for all $j=1,\dots,J$. To keep notation manageable, we explicitly denote the estimator used in (2.5) by $\hat{\sigma}^M_j$ only in this step but in almost all other parts of this Appendix we use the generic notation $\hat{\sigma}_j$.

For each $j=1,\dots,R_1$ such that

$$\pi^*_{1,j}=\pi^*_{1,j+R_1}=0, \tag{D.10}$$

where π^*_1 is defined in (D.5), let

$$\tilde{\mu}_j=\begin{cases}1 & \text{if } \gamma_{1,P_n,j}(\theta_n)=0=\gamma_{1,P_n,j+R_1}(\theta_n),\\ \frac{\gamma_{1,P_n,j+R_1}(\theta_n)(1+\eta_{n,j+R_1}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))}{\gamma_{1,P_n,j+R_1}(\theta_n)(1+\eta_{n,j+R_1}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))+\gamma_{1,P_n,j}(\theta_n)(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))} & \text{otherwise,}\end{cases} \tag{D.11}$$

$$\tilde{\mu}_{j+R_1}=\begin{cases}0 & \text{if } \gamma_{1,P_n,j}(\theta_n)=0=\gamma_{1,P_n,j+R_1}(\theta_n),\\ \frac{\gamma_{1,P_n,j}(\theta_n)(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))}{\gamma_{1,P_n,j+R_1}(\theta_n)(1+\eta_{n,j+R_1}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))+\gamma_{1,P_n,j}(\theta_n)(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))} & \text{otherwise,}\end{cases} \tag{D.12}$$

For each $j=1,\dots,R_1$, replace the constraint indexed by j, that is

$$\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}\leq\hat{c}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}), \tag{D.13}$$

with the following weighted sum of the paired inequalities

$$\tilde{\mu}_j\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}-\tilde{\mu}_{j+R_1}\frac{\sqrt{n}\bar{m}_{j+R_1,n}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j+R_1}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}\leq\hat{c}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}), \tag{D.14}$$

and for each $j=1,\dots,R_1$, replace the constraint indexed by $j+R_1$, that is

$$\frac{\sqrt{n}\bar{m}_{j+R_1,n}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j+R_1}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})}\leq\hat{c}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}), \tag{D.15}$$

with

$$-\tilde{\mu}_j \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} + \tilde{\mu}_{j+R_1} \frac{\sqrt{n}\bar{m}_{j+R_1,n}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j+R_1}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \le \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \tag{D.16}$$

It then follows from Assumption 4.3-(II) that these replacements are conservative because

$$\frac{\bar{m}_{j+R_1,n}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j+R_1}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \le -\frac{\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}^M_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})},$$

and therefore (D.14) implies (D.13) and (D.16) implies (D.15).

Step 3. Next, we make the following comparisons:

$$\pi^*_{1,j} = 0 \Rightarrow \pi^*_{1,j} \ge \sqrt{n}\gamma_{1,P_n,j}(\theta_n), \tag{D.17}$$

$$\pi^*_{1,j} = -\infty \Rightarrow \sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to -\infty. \tag{D.18}$$

For any constraint j for which $\pi^*_{1,j} = 0$, (D.17) yields that replacing $\sqrt{n}\gamma_{1,P_n,j}(\theta_n)$ in (D.9) with $\pi^*_{1,j}$ introduces a conservative distortion. Under Assumption 4.3-(II), for any j such that (D.10) holds, the substitutions in (D.14) and (D.16) yield $\tilde{\mu}_j\sqrt{n}\gamma_{1,P_n,j}(\theta_n)(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) - \tilde{\mu}_{j+R_1}\sqrt{n}\gamma_{1,P_n,j+R_1}(\theta_n)(1 + \eta_{n,j+R_1}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) = 0$, and therefore replacing this term with $\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1}$ is inconsequential.

For any j for which $\pi^*_{1,j} = -\infty$, (D.18) yields that for n large enough, $\sqrt{n}\gamma_{1,P_n,j}(\theta_n)$ can be replaced with $\pi^*_{1,j}$. To see this, note that by the Cauchy-Schwarz inequality, Assumption 4.4 (i)-(ii), and $\lambda \in B^d_{n,\rho}$, it follows that

$$\rho D_{P_n,j}(\bar{\theta}_n)\lambda \le \rho\sqrt{d}(\|D_{P_n,j}(\bar{\theta}_n) - D_{P_n,j}(\theta_n)\| + \|D_{P_n,j}(\theta_n)\|) \le \rho\sqrt{d}(\rho M/\sqrt{n} + \bar{M}), \tag{D.19}$$

where $\bar{M}$ and M are as defined in Assumption 4.4-(i) and (ii) respectively, and we used that $\bar{\theta}_n$ lies component-wise between θ_n and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$. Using that $\mathbb{G}_{n,j}$ is asymptotically tight by Assumption 4.5, we have that for any $\tau > 0$, there exists a $T > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \ge N_1$,

$$P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le 0,\ \forall \lambda \in B^d_{n,\rho}\right) > 1 - \tau/2. \tag{D.20}$$

To see this, note that $\pi^*_{ij} = -\infty$ if and only if $\lim_{n\to\infty}\frac{\sqrt{n}}{\kappa_n}\gamma_{1P_nj}(\theta_n) = \pi_{1j} \in [-\infty, 0)$. Suppose first that $\pi_{1j} > -\infty$. Then for all $\epsilon > 0$ there exists $N_2 \in \mathbb{N}$ such that $\left|\frac{\sqrt{n}}{\kappa_n}\gamma_{1P_nj}(\theta_n) - \pi_{1j}\right| \le \epsilon$, for all $n \ge N_2$. Choose $\epsilon > 0$ such that $\pi_{1j} + \epsilon < 0$. Let $N = \max\{N_1, N_2\}$. Then we have

$$\begin{aligned}
&P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le 0,\ \forall \lambda \in B^d_{n,\rho}\right) \\
\ge\ & P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\{T + \rho(\bar{M} + \rho M/\sqrt{n}) + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le 0 \cap \max_{j:\pi^*_{1,j}=-\infty}\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \le T\right) \\
\ge\ & P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\{T + \rho(\bar{M} + \rho M/\sqrt{n}) + \kappa_n(\pi_{1j} + \epsilon)\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le 0 \cap \max_{j:\pi^*_{1,j}=-\infty}\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \le T\right) \\
=\ & P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\left\{\frac{T}{\kappa_n} + \frac{\rho}{\kappa_n}(\bar{M} + \rho M/\sqrt{n}) + (\pi_{1j} + \epsilon)\right\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \le 0 \cap \max_{j:\pi^*_{1,j}=-\infty}\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \le T\right) \\
=\ & P_n\left(\max_{j:\pi^*_{1,j}=-\infty}\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \le T\right) > 1 - \tau/2,\ \forall n \ge N.
\end{aligned}$$

If $\pi_{1j} = -\infty$ the same argument applies a fortiori. We therefore have that for $n \geq N$,

$$P_n\Bigg(\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\} \leq 0$$

$$\leq \left\{\begin{array}{c}\sup_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\}\Bigg) \tag{D.21}$$

$$\geq P_n\Bigg(\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\} \leq 0$$

$$\leq \left\{\begin{array}{c}\sup_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\}\Bigg) - \tau/2. \tag{D.22}$$

Since the choice of τ is arbitrary, the limit of the term in (D.21) is not smaller than the limit of the first term in (D.22). Hence, we continue arguing for the event whose probability is evaluated in (D.22).

Finally, by definition $\hat{c}_n(\cdot) \geq 0$ and therefore $\inf_{\lambda\in B^d_{n,\rho}} \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}})$ exists. Therefore, the event whose probability is evaluated in (D.22) is implied by the event

$$\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \inf_{\lambda\in B^d_{n,\rho}} \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\} \leq 0$$

$$\leq \left\{\begin{array}{c}\sup_\lambda p'\lambda\\ s.t. \lambda \in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})) \leq \inf_{\lambda\in B^d_{n,\rho}} \hat{c}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j\end{array}\right\} \tag{D.23}$$

For each $\lambda \in \mathbb{R}^d$, define

$$u_{n,j,\theta_n}(\lambda) \equiv \{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})), \tag{D.24}$$

where under Assumption 4.3-(II) when $\pi^*_{1,j} = 0$ and $\pi^*_{1,j+R_1} = 0$ the substitutions of equation (D.13) with equation (D.14) and of equation (D.15) with equation (D.16) have been performed. Let

$$U_n(\theta_n, c) \equiv \left\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u_{n,j,\theta_n}(\lambda) \leq c, \ \forall j = 1, \ldots, J\right\}, \tag{D.25}$$

and define

$$\hat{c}_{n,\rho} \equiv \inf_{\lambda\in B^d_{n,\rho}} \hat{c}_n(\theta + \frac{\lambda\rho}{\sqrt{n}}). \tag{D.26}$$

Then by (D.23) and the definition of U_n, we obtain

$$P_n(p'\theta_n \in CI_n) \geq P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}) \neq \emptyset\right). \tag{D.27}$$

By passing to a further subsequence, we may assume that

$$D_{P_n}(\theta_n) \to D, \tag{D.28}$$

for some $J \times d$ matrix D such that $\|D\| \leq M$ and $\Omega_{P_n} \xrightarrow{u} \Omega$ for some correlation matrix Ω. By Lemma 2 in Andrews and Guggenberger (2009) and Assumption 4.5 (i), uniformly in $\lambda \in B^d$, $\mathbb{G}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \xrightarrow{d} \mathbb{Z}$ for a normal random vector with the correlation matrix Ω. By Lemma E.1,

$$\liminf_{n\to\infty} P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}) \neq \emptyset\right) \geq 1 - \alpha. \tag{D.29}$$

The conclusion of the theorem then follows from (D.2), (D.3), (D.27), and (D.29). □

Proof of Theorem 4.2

The argument of proof is the same as for Theorem 4.1, with the following modification. Take (P_n, θ_n) as defined following equation (D.4). Then $f(\theta_n)$ is covered when

$$\begin{aligned}
&\begin{Bmatrix} \inf f(\vartheta) \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq \hat{c}_n^f(\vartheta), \forall j \end{Bmatrix} \leq f(\theta_n) \leq \begin{Bmatrix} \sup f(\vartheta) \\ s.t.\, \vartheta \in \Theta, \quad \frac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq \hat{c}_n^f(\vartheta), \forall j \end{Bmatrix} \\
\Leftrightarrow &\begin{Bmatrix} \inf_\lambda \nabla f(\tilde{\theta}_n)\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \leq \hat{c}_n^f(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix} \leq 0 \\
&\leq \begin{Bmatrix} \sup_\lambda \nabla f(\tilde{\theta}_n)\lambda \\ s.t. \lambda \in \frac{\sqrt{n}}{\rho}(\Theta - \theta_n), \quad \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \leq \hat{c}_n^f(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \forall j \end{Bmatrix},
\end{aligned}$$

where we took a mean value expansion yielding

$$f(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) = f(\theta_n) + \frac{\rho}{\sqrt{n}}\nabla f(\tilde{\theta}_n)\lambda, \tag{D.30}$$

for $\tilde{\theta}_n$ a mean value that lies componentwise between θ_n and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$, and we used that the sign of the last term in (D.30) is the same as the sign of $\nabla f(\tilde{\theta}_n)\lambda$. With the objective function in (D.30) so redefined, all expression in the proof of Theorem 4.1 up to (D.24) continue to be valid. We can then redefine the set $U_n(\theta_n, c)$ in (D.25) as

$$U_n(\theta_n, c) \equiv \left\{\lambda \in B^d_{n,\rho} : \|\nabla f(\tilde{\theta}_n)\|^{-1}\nabla f(\tilde{\theta}_n)\lambda = 0 \cap u_{n,j,\theta_n}(\lambda) \leq c, \ \forall j = 1, \dots, J\right\}.$$

Replace p' with $\|\nabla f(\tilde{\theta}_n)\|^{-1}\nabla f(\tilde{\theta}_n)$ in all expressions involving the set $U_n(\theta_n, \hat{c}^f_{n,\rho}(\theta_n))$, and replace p' with $\|\nabla f(\theta_n)'\|^{-1}\nabla f(\theta_n')$ in all expressions for the sets $V_n^I(\theta_n', \hat{c}_n^f(\theta_n'))$, and in all the almost sure representation counterparts of these sets. Observe that we can select a convergent subsequence from $\{\|\nabla f(\theta_n)'\|^{-1}\nabla f(\theta_n')\}$ that converges to some p in the unit sphere, so that the form of $\mathfrak{W}(c_{\pi^*})$ in (E.17) is unchanged. This yields the result, noting that by the assumption $\|\nabla f(\tilde{\theta}_n) - \nabla f(\theta_n')\| = O_{\mathcal{P}}(\rho/\sqrt{n})$ □

D.2.2 A High Level Condition Replacing Assumption 4.3 and the ρ-Box Constraints

Next, we consider an assumption which is composed of two parts. The first part aims at informally mimicking Assumption A.2 in Bugni, Canay, and Shi (2017) and replaces Assumption 4.3. The second part replaces the use of the ρ-box constraints. Below, for a given set $A \subset \mathbb{R}^d$, let $\|A\|_H = \sup_{a\in A}\|a\|$ denote its Hausdorff norm.

Assumption D.1: *Consider any sequence* $\{P_n, \theta_n\} \in \{(P,\theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ *such that*

$$\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]},\ j = 1,\dots,J,$$
$$\Omega_{P_n} \overset{u}{\to} \Omega,$$
$$D_{P_n}(\theta_n) \to D.$$

Let $\pi_{1j}^* = 0$ *if* $\pi_{1j} = 0$ *and* $\pi_{1j}^* = -\infty$ *if* $\pi_{1j} < 0$. *Let* $\mathbb{Z}$ *be a Gaussian process with covariance kernel* Ω. *Let*

$$\mathfrak{w}_j(\lambda) \equiv \mathbb{Z}_j + \rho D_j \lambda + \pi_{1,j}^*. \tag{D.31}$$

(I) Let

$$\mathfrak{W}(c) \equiv \left\{\lambda \in \mathfrak{B}_\rho^d : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c,\ \forall j = 1,\dots,J\right\}, \tag{D.32}$$
$$c_{\pi^*} \equiv \inf\{c \in \mathbb{R}_+ : \Pr(\mathfrak{W}(c) \neq \emptyset) \ge 1-\alpha\}. \tag{D.33}$$

Then:

(a) If $c_{\pi^*} > 0$, $\Pr(\mathfrak{W}(c) \neq \emptyset)$ *is continuous and strictly increasing at* $c = c_{\pi^*}$.

(b) If $c_{\pi^*} = 0$, $\liminf_{n\to\infty} P_n(U_n(\theta_n, 0) \neq \emptyset) \ge 1-\alpha$, *where* $U_n(\theta_n, c),\ c \ge 0$ *is as in* (D.25).

(II) Let

$$\bar{\mathfrak{W}}(c) \equiv \left\{\lambda \in \mathbb{R}^d : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c,\ \forall j = 1,\dots,J\right\},$$

which differs from (D.32) *by not constraining* λ *to* $\mathfrak{B}_\rho^d$, *and let* $\bar{c} \equiv \Phi^{-1}(1-\alpha/J)$ *denote the asymptotic Bonferroni critical value. Then for every* $\eta > 0$ *there exists* $M_\eta < \infty$ *s.t.* $\Pr(\|\bar{\mathfrak{W}}(\bar{c})\|_H > M_\eta) \le \eta$.

D.2.3 Proof of Theorem 4.1 with High Level Assumption D.1-I Replacing Assumption 4.3, and Dropping the ρ-Box Constraints Under Assumption D.1-II

Lemma D.1: *Suppose that Assumption 4.1, 4.2, 4.4 and 4.5 hold.*

(I) Let also Assumption D.1-I hold. Let $0 < \alpha < 1/2$. *Then,*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n) \ge 1-\alpha.$$

(II) Let also Assumption D.1-II and either Assumption 4.3 or D.1-I hold. Let $\hat{c}_n = \inf\{c \in \mathbb{R}_+ : P^*(\{\Lambda_n^b(\theta, +\infty, c) \cap \{p'\lambda = 0\}\} \neq \emptyset) \ge 1-\alpha\}$, *where* Λ_n^b *is defined in equation* (3.1) *and* $CI_n \equiv [-s(-p, \mathcal{C}_n(\hat{c}_n)), s(p, \mathcal{C}_n(\hat{c}_n))]$ *with* $s(q, \mathcal{C}_n(\hat{c}_n)), q \in \{p, -p\}$ *defined in equation* (2.5). *Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in CI_n) \ge 1-\alpha.$$

Proof. We establish each part of the Lemma separately.

Part I. This part of the lemma replaces Assumptions 4.3 with Assumption D.1-I. Hence we establish the result by showing that all claims that were made under Assumption 4.3 remain valid under Assumption D.1-I. We proceed in steps.

Step 1. Revisiting the proof of Lemma E.6, equation (E.133).

Let $\mathcal{J}^*$ be as defined in (E.29). If $\mathcal{J}^* = \emptyset$ we immediately have that Lemma E.6 continues to hold. Hence we assume that $\mathcal{J}^* \neq \emptyset$. To keep the notation simple, below we argue as if all $j = 1,\dots,J$ belong to $\mathcal{J}^*$.

Consider the case that $c_{\pi^*} > 0$. For some $c_{\pi^*} > \delta > 0$, let

$$\mathfrak{W}(c-\delta) \equiv \{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \leq c - \delta,\ \forall j = 1, \ldots, J\}, \tag{D.34}$$

where we emphasize that the set $\mathfrak{W}(c-\delta)$ is obtained by a δ-contraction of all constraints, including those indexed by $j = J_1 + 1, \ldots, J$. By Assumption D.1-I, for any $\eta > 0$ there exists a δ such that

$$\eta \geq |\Pr(\mathfrak{W}(c_{\pi^*}) \neq \emptyset) - \Pr(\mathfrak{W}(c_{\pi^*} - \delta) \neq \emptyset)| = \Pr(\{\mathfrak{W}(c_{\pi^*}) \neq \emptyset\} \cap \{\mathfrak{W}(c_{\pi^*} - \delta) = \emptyset\}),$$
$$\eta \geq |\Pr(\mathfrak{W}(c_{\pi^*} + \delta) \neq \emptyset) - \Pr(\mathfrak{W}(c_{\pi^*}) \neq \emptyset)| = \Pr(\{\mathfrak{W}(c_{\pi^*} + \delta) \neq \emptyset\} \cap \{\mathfrak{W}(c_{\pi^*}) = \emptyset\}).$$

The result follows.

Step 2. Revisiting the proof of Lemma E.2.

Case 1 of Lemma E.2 is unaltered. Case 2 of Lemma E.2 follows from the same argument as used in Case 1 of Lemma E.2, because under Assumption D.1-I as shown in step 1 of this proof all inequalities are tightened. In Case 3 of Lemma E.2 the result in (D.29) holds automatically by Assumption D.1-I-(ii). (As a remark, Lemmas E.7-E.8 are no longer needed to establish Lemma E.2.)

Step 3. Revisiting the proof of Lemma E.3. Under Assumption D.1 we do not need to merge paired inequalities. Hence, part (iii) of Lemma E.3 holds automatically because $\varphi^*_j(\xi) \leq \varphi_j(\xi)$ for any j and ξ. We are left to establish parts (i) and (ii) of Lemma E.3. These follow immediately, because Lemma E.6 remains valid as shown in step 1 and by Assumption D.1-I, $\Pr(\mathfrak{W}(c) \neq \emptyset)$ is strictly increasing at $c = c_{\pi^*}$ if $c_{\pi^*} > 0$. (As a remark, Lemma E.9 is no longer needed to establish Lemma E.3.)

In summary, the desired result follows by applying Lemma E.1 in the proof of Theorem 4.1 as Lemmas E.2, E.3 and E.6 remain valid, Lemmas E.4, E.5, E.10 and the Lemmas in Appendix E.3 are unaffected, and Lemmas E.7, E.8, E.9 are no longer needed.

Part II. This is established by adapting the proof of Theorem 4.1 as follows:

In the main proof, we pass to an a.s. representation early on, so that $\mathfrak{W}$ realizes jointly with other random variables (we denote almost sure representations adding a superscript "*" on the original variable). At the same time, we entirely drop ρ. This means that algebraic expressions, e.g. in the main proof, simplify as if $\rho = 1$, but it also removes any constraints along the lines of $\lambda \in B^d_{n,\rho}$ in equation (D.9). Indeed, (D.9) is replaced by:

$$\cdots \Leftarrow \left\{ \begin{array}{c} \inf_\lambda p'\lambda \\ s.t. \lambda \in \bar{\mathfrak{W}}^*(\bar{c}), \\ \{\mathbb{G}^*_{n,j}(\theta_n + \lambda/\sqrt{n}) + D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \lambda/\sqrt{n})) \leq \hat{c}_n(\theta_n + \lambda/\sqrt{n}), \forall j \end{array} \right\} \leq 0$$
$$\leq \left\{ \begin{array}{c} \sup_\lambda p'\lambda \\ s.t. \lambda \in \bar{\mathfrak{W}}^*(\bar{c}), \\ \{\mathbb{G}^*_{n,j}(\theta_n + \lambda/\sqrt{n}) + D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1 + \eta_{n,j}(\theta_n + \lambda/\sqrt{n})) \leq \hat{c}_n(\theta_n + \lambda/\sqrt{n}), \forall j \end{array} \right\},$$

yielding a new definition of the set U^*_n as

$$U^*_n(\theta_n, c) \equiv \{\lambda \in \bar{\mathfrak{W}}^*(\bar{c}) : p'\lambda = 0 \cap u^*_{n,j,\theta_n}(\lambda) \leq c,\ \forall j = 1, \ldots, J\}.$$

Subsequent uses of ρ in the main proof use that $\|\lambda\| \leq \sqrt{d}\rho = O_{\mathcal{P}}(1)$. For example, consider the argument following equation (E.30) or the argument just preceding equation (D.29), and so on. All these continue to go through because $\bar{\mathfrak{W}}^*(\bar{c}) = O(1)$ by assumption.

Similar uses occur in Lemma E.1. The next major adaptation is that in (E.27) and (E.28): we again drop ρ

but nominally introduce the constraint that $\lambda \in \bar{\mathfrak{W}}^*(\bar{c})$. However, for $c \leq \bar{c}$, this condition cannot constrain $\mathfrak{W}^*(c)$, and so we can as well drop it: The modified $\mathfrak{W}^*(c)$ equals $\bar{\mathfrak{W}}^*(c)$.

Next we argue that Lemma E.7 continues to hold, now claimed for $\bar{\mathfrak{W}}^*$. To verify that this is the case, replace B^d with $\bar{\mathfrak{W}}(\bar{c})$ throughout in Lemma E.7. This requires straightforward adaptation of algebra as $\bar{\mathfrak{W}}(\bar{c})$ is only stochastically and not deterministically bounded.

Finally, in Lemma E.3 we remove the ρ-constraint from V_n^b and V_n^I without replacement, and note that the lemma is now claimed for $\theta'_n \in \theta + \|\bar{\mathfrak{W}}(\bar{c})\|_H/\sqrt{n}B^d$. Recall that in the lemma the a.s. representation of a set A is denoted by $\tilde{A}$, and with some abuse of notation let the a.s. representation of $\bar{\mathfrak{W}}$ be denoted $\widetilde{\bar{\mathfrak{W}}}$. Now we compare $\tilde{V}_n^b$ and $\tilde{V}_n^I$ with $\widetilde{\bar{\mathfrak{W}}}$. To ensure that λ is uniformly stochastically bounded in expressions like (E.95), we verify that the modified $\tilde{V}_n^b$ and $\tilde{V}_n^I$ inherit the property in Assumption D.1-II. To see this, fix any unit vector $t \perp p$ and notice that any $t = \lambda/\|\lambda\|$ for $\lambda \in \widetilde{\bar{\mathfrak{W}}}(c)$ or for $\lambda \in \tilde{V}_n^b(\theta'_n, c)$ or for $\lambda \in \tilde{V}_n^I(\theta'_n, c)$, $0 < c \leq \bar{c}$, satisfies this condition. By Assumption D.1-II and the Cauchy-Schwarz inequality, $\max_{\lambda \in \widetilde{\bar{\mathfrak{W}}}(c)} t'\lambda = O(1)$ for any $c \leq \bar{c}$. Since the value of this program is necessarily attained by a basic solution whose associated gradients span t, it must be the case that such solution is itself $O(1)$. Formally, let C be the index set characterizing the solution, $\mathbb{Z}_i^C$ be the vector of realizations $\mathbb{Z}_i^j$ corresponding to $j \in C$, and $K^C(\theta'_n)$ the matrix that stacks the corresponding gradients; then $(K^C(\theta'_n))^{-1}(\bar{c}\mathbf{1} - \mathbb{Z}_i^C) = O(1)$. By Lemma E.7 and the fact that $\hat{D}_n(\theta'_n) \xrightarrow{P} D$ by Assumption 4.4, we then also have that $(\hat{K}^C(\theta'_n))^{-1}(\bar{c}\mathbf{1} - \mathbb{G}_{n,j}^b) = O_{\mathcal{P}}(1)$, and so for $c \leq \bar{c}$, V^b is bounded in this same direction. It follows that, by similar reasoning to the preceding paragraph, the comparison between $V_n^I(\theta'_n, c)$ and $\bar{\mathfrak{W}}(c)$ in Lemma E.3 goes through. □

D.3 Proof of Theorems 4.3 and 4.4

D.3.1 Assumptions in Pakes, Porter, Ho, and Ishii (2011), Chernozhukov, Hong, and Tamer (2007), and Bugni, Canay, and Shi (2017) That Allow for Simplifications of the Method

We analyze calibrated projection under assumptions that are more stringent than for Theorem 4.1. The reward is considerable computational simplification and, in some cases, removal of a tuning parameter. The additional assumptions have been used in the related literature. Their logical relation to each other and to explicit constraint qualifications is further analyzed in Kaido, Molinari, and Stoye (2017). For our purposes in this paper, we just state without proof that, given Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5, all assumptions below, including the minorant assumptions attributed to other papers, are implied by assumptions in Pakes, Porter, Ho, and Ishii (2011); hence, all results reported below apply under the Pakes, Porter, Ho, and Ishii (2011) assumptions.[5]

For $\theta \in \partial\Theta_I(P)$, denote by $\mathcal{J}(P,\theta)$ the set of inequalities j s.t. $E_P(m_j(X_i,\theta)) = 0$. Denote by $\mathcal{N}(P,\theta)$ the positive span of $(D_{P,j})_{j \in \mathcal{J}(P,\theta)}$ and by $\mathcal{T}(P,\theta) = \{t : D'_{P,j}t \leq 0, j \in \mathcal{J}(P,\theta)\}$ the corresponding dual cone. (These are the normal and tangent cones of $\Theta_I(P)$ at θ.) For a given $p \in \mathbb{R}^d : \|p\| = 1$, let $s(p, \Theta_I(P)) = \max_{\theta \in \Theta_I(P)} p'\theta$ and $H(p, \Theta_I(P)) \equiv \arg\max_{\theta \in \Theta_I(P)} p'\theta$.

Assumption D.2 (A weakening of Assumption 4(a) in Pakes, Porter, Ho, and Ishii (2011)): *There is a class of DGPs $\mathcal{Q} \subset \mathcal{P}$ such that any $P \in \mathcal{Q}$ satisfies the following conditions:*

[5] Our own assumptions meaningfully exceed those of Pakes, Porter, Ho, and Ishii (2011) only through Assumption 4.3. The absence of such an assumption in Pakes, Porter, Ho, and Ishii (2011) is actually an oversight, and ours or a similar assumption must be added for their Theorem 2 to hold.

1. There exists a (universal) $\varepsilon_D > 0$ s.t.

$$\min_{\theta \in H(p,\Theta_I(P))} \min_{\|t\|=1} \max_{\substack{j \in \{1,\dots,J\}: \\ E_P(m_j(X_i,\theta)/\sigma_j(\theta)) > -\varepsilon_D}} t' D_{P,j}(\theta) < -\varepsilon_D.$$

2. There exists a (universal) $\varepsilon_D > 0$ s.t.

$$\max_{\theta \in H(p,\Theta_I(P))} \min_{\|t\|=1} \max_{\substack{j \in \{1,\dots,J\}: \\ E_P(m_j(X_i,\theta)/\sigma_j(\theta)) > -\varepsilon_D}} t' D_{P,j}(\theta) < -\varepsilon_D.$$

There are two layers to these assumptions. First, they say that from some support point (part (1)) or all support points (part (2)), there are directions that point uniformly inside $\Theta_I(P)$ in the sense of all moment inequalities decreasing in value. The obvious counterexample would be an extremely pointy corner (a "spike").

In addition, the assumptions apply to "tightened" tangent cones that use all inequalities which are almost binding, where "almost" is operationalized with the small but positive constant ε_D. Together with smoothness of moment conditions, this implies that, by moving a small (but boundedly nonzero) distance in the direction of steepest descent from the support point, one can find a point θ at which $\max_j E_P\left(m_j(X_i,\theta)/\sigma_j(\theta)\right)$ is boundedly negative. This implies that the sample analog of $\Theta_I(P)$ is nonempty with probability approaching 1 (the proof in Appendix D.3.2 includes a formal version of this argument). In particular, it implies that a vestige of the "degeneracy" assumption in Chernozhukov, Hong, and Tamer (2007) is imposed. Some invocations of the assumption strictly speaking only use one of the two features (again, see Kaido, Molinari, and Stoye (2017) for details), but we do not disentangle them here. Note, however, that the second implication renders the assumption implausible whenever the sample analog of $\Theta_I(P)$ is empty, an empirically frequent occurrence.

Next, consider:

Assumption D.3 (Linear Minorant – Chernozhukov, Hong, and Tamer (2007) display (4.5)): *There exist universal constants $C, \delta > 0$ and a class of DGPs $\mathcal{Q} \subset \mathcal{P}$ such that for each $P \in \mathcal{Q}$,*

$$\max_{j=1,\dots,J} E_P\left(m_j(X_i,\theta)/\sigma_j(\theta)\right) \geq C \min\left\{\delta, d\left(\theta, \Theta_I(P)\right)\right\}.$$

Assumption D.4 (Linear Minorant Along Support Plane – Bugni, Canay, and Shi (2017) Assumption A3(a)): *There exist universal constants $C, \delta > 0$ and a class of DGPs $\mathcal{Q} \subset \mathcal{P}$ such that for each $P \in \mathcal{Q}$ and for each $q \in \{p, -p\}$,*

$$\max_{j=1,\dots,J} E_P\left(m_j(X_i,\theta)/\sigma_j(\theta)\right) \geq C \min\left\{\delta, d\left(\theta, H(q,\Theta_I(P))\right)\right\}$$

for all θ with $q'\theta = s(q, \Theta_I(P))$.

These assumptions are lifted from the cited papers. In the original papers, they are polynomial minorant conditions: The minima are raised to some power χ. However, for our setting and criterion function, the special case $\chi = 1$ applies. Note also that Assumption D.4 is closely analogous to Assumption D.3 but imposes the minorant condition on the "null restricted model in which the parameter space is restricted to the true supporting hyperplane of $\Theta_I(P)$. It is easy to see that the assumptions are logically independent.

A further strengthening of assumptions is:

Assumption D.5 (A Weakening of Assumption 3 in Pakes, Porter, Ho, and Ishii (2011)): *There exists a universal constant $\overline{\delta} > 0$ and a class of DGPs $\mathcal{Q} \subset \mathcal{P}$ such that for any $P \in \mathcal{Q}$ and for each $q \in \{p, -p\}$ and any $\theta \in H(q, \Theta_I(P))$, $\mathcal{T}(\theta) \subseteq \{t : q't/\|t\| \leq -\overline{\delta}\}$.*

Note the implication that $\mathcal{T}(P,\theta)$ is uniformly pointy. The assumption is weaker than in Pakes, Porter, Ho, and Ishii (2011) because they also assume $\Theta_I(P) \subseteq \mathcal{T}(\theta)$ and separately (although it is also an implication) that $H(p, \Theta_I(P))$ is a singleton.

Our final assumption gives a further strengthening by requiring the support set in direction of projection to be a singleton:

Assumption D.6 (Assumption 1 in Pakes, Porter, Ho, and Ishii (2011)): *There is a class of DGPs $\mathcal{Q} \subset \mathcal{P}$ such that for any $P \in \mathcal{Q}$ and $q \in \{p, -p\}$, $H(q, \Theta_I(P))$ is a singleton. (Its sole element will be denoted θ_q^* below.)*

D.3.2 Proof of Theorem 4.3: Simplifications for Calibrated Projection

Part I

Let θ_p^* attain the outer minimum in Assumption D.2-1, let t^* attain the inner minimum given θ_p^*, and consider any $\eta \leq \varepsilon_D/2M$, where ε_D is from Assumption D.2-1 and M is from Assumption 4.4(ii). Then a Mean Value Theorem yields

$$\begin{aligned}
\frac{E_P(m_j(X_i, \theta_p^* + \eta t^*))}{\sigma_{P,j}(\theta_p^* + \eta t^*)} &= \frac{E_P(m_j(X_i, \theta_p^*))}{\sigma_{P,j}(\theta_p^*)} + \eta D_{P_j}(\bar{\theta}) t^* \\
&\leq 0 + \eta(\eta M - \varepsilon_D) \\
\Longrightarrow \max_j \frac{E_P(m_j(X_i, \theta_p^* + \eta t^*))}{\sigma_{P,j}(\theta_p^* + \eta t^*)} &\leq -\eta \varepsilon_D/2. \qquad \text{(D.35)}
\end{aligned}$$

This will be used later but also implies

$$\begin{aligned}
&\max_j \frac{E_P(m_j(X_i, \theta_p^* + t^* \varepsilon_D/2M))}{\sigma_{P,j}(\theta_p^* + t^* \varepsilon_D/2M)} \leq -\varepsilon_D^2/4M < 0 \qquad \text{(D.36)} \\
\Longrightarrow \quad & P\left(\max_j \frac{\bar{m}_j(X_i, \theta_p^* + t^* \varepsilon_D/2M)}{\hat{\sigma}_j(\theta_p^* + t^* \varepsilon_D/2M))} < 0 \right) \to 1 \\
\Longrightarrow \quad & P(\theta_p^* + t^* \varepsilon_D/2M \in CI_n) \to 1
\end{aligned}$$

uniformly in $\mathcal{Q}$. Hence, noncoverage risk for any $\gamma \in \left[-s(-p, \Theta_I(P)), p'(\theta_p^* + t^* \varepsilon_D/2M)\right]$ is entirely driven by the possibility that CI_n is too high, and conversely for $\gamma \in \left[p'(\theta_p^* + t^* \varepsilon_D/2M, s(p, \Theta_I(P)))\right]$. As these noncoverage risks are monotonic in γ, the simplification is justified. $\square$

Part II

Note first that, as an immediate implication of D.36, the event that $\min_{\theta \in \Theta} \max_j |\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_+ = 0$, hence this value is attained on $\hat{\Theta}_I$, occurs w.p.a. 1 uniformly in $\mathcal{Q}$.

Next, we show that $\sqrt{n}(s(p, \hat{\Theta}_I) - s(p, \Theta_I(P))) = O_{\mathcal{Q}}(1)$. Define

$$\mathcal{C}(-\varepsilon) = \left\{ \theta \in \Theta : \max_{j=1,\dots,J} E_P\left(m_j(X_i, \theta)/\sigma_{P,j}(\theta)\right) \leq -\varepsilon \right\}.$$

Note that in this notation, $\Theta_I(P) = \mathcal{C}(0)$. By (D.36) and because $\mathcal{C}(-\varepsilon)$ is closed by assumptions on m_j, we have that $H(p, \mathcal{C}(-\varepsilon))$ is nonempty for $\varepsilon \in [0, \varepsilon_D^2/4M]$. Next, consider any $\eta \leq \varepsilon_D/2M$, then $p'(\theta_p^* + \eta t^*) \geq p'\theta_p^* - \eta$, which together with (D.35) implies

$$s(p, \mathcal{C}(-\eta \varepsilon_D/2)) - s(p, \Theta_I) \geq -\eta.$$

Set $\varepsilon = \eta \varepsilon_D/2$, then equivalently we find that for $\varepsilon \leq \varepsilon_D^2/4M$, $s(p, \mathcal{C}(-\varepsilon) - s(p, \Theta_I)) \geq -2\varepsilon/\varepsilon_D$. Next, we have that uniformly over $\theta \in \cup_{\varepsilon\in[0,\varepsilon_D^2/4M]} H(p, \mathcal{C}(-\varepsilon))$,

$$\begin{aligned}
\sqrt{n} \max_j |\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_+ &= \max_j \left\{ (1-\eta_{n,j}(\theta)) \left| \mathbb{G}_{n,j}(\theta) + \sqrt{n} E_P \left(m_j(X_i,\theta)/\sigma_{P,j}(\theta) \right) \right|_+ \right\} \\
&\leq \sum_j (1-\eta_{n,j}(\theta)) \left| \mathbb{G}_{n,j}(\theta) + \sqrt{n} E_P \left(m_j(X_i,\theta)/\sigma_{P,j}(\theta) \right) \right|_+ \\
&\leq J(1+o_{\mathcal{Q}}(1)) \left| O_{\mathcal{Q}}(1) - \sqrt{n}\varepsilon \right|_+ ,
\end{aligned}$$

so in analogy to CHT (Theorem 4.2, step 1 of proof) we find $\sqrt{n}|s(p,\hat{\Theta}_I) - s(p,\Theta_I)|_- = O_{\mathcal{Q}}(1)$. On the other hand, from Assumption D.3 we have that uniformly over $\theta \in \Theta$,

$$\begin{aligned}
\sqrt{n} \max_j |\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_+ &= \max_j \left\{ (1-\eta_{n,j}(\theta)) \left| \mathbb{G}_{n,j}(\theta) + \sqrt{n} E_P \left(m_j(X_i,\theta)/\sigma_{P,j}(\theta) \right) \right|_+ \right\} \\
&\geq \frac{1}{J} \sum_{j=1}^J (1-\eta_{n,j}(\theta)) \left| \mathbb{G}_{n,j}(\theta) + \sqrt{n} E_P \left(m_j(X_i,\theta)/\sigma_{P,j}(\theta) \right) \right|_+ \\
&\geq \frac{1}{J} \sum_{j=1}^J (1+o_{\mathcal{Q}}(1)) \left| O_{\mathcal{Q}}(1) + \sqrt{n} C \min\left\{\delta, d(\theta, \Theta_I(P))\right\} \right|_+ ,
\end{aligned}$$

hence $\sqrt{n}|s(p,\hat{\Theta}_I) - s(p,\Theta_I(P))|_+ = O_{\mathcal{Q}}(1)$.

We next argue that $d(\hat{\theta}_p, H(p,\Theta_I(P))) = O_{\mathcal{Q}}(n^{-1/2})$ (the proof for $d(\hat{\theta}_{-p}, H(-p,\Theta_I(P)))$ is identical). To do so, let $\hat{k} \equiv s(p,\Theta_I(P)) - s(p,\hat{\Theta}_I)$ and define $\tilde{\theta} = \hat{\theta}_p + \hat{k}p$, noting that $p'\tilde{\theta} = s(p,\Theta_I(P))$ by construction and so Assumption D.4 applies to $\tilde{\theta}$. Let $\bar{\theta} \in H(p,\Theta_I(P))$ be such that $d(\tilde{\theta}, H(p,\Theta_I(P))) = \|\bar{\theta} - \tilde{\theta}\|$, then

$$d\left(\hat{\theta}_p, H(p,\Theta_I(P))\right) \leq d\left(\hat{\theta}_p, \tilde{\theta}\right) + d\left(\tilde{\theta}, H(p,\Theta_I(P))\right) \leq \left\|\hat{\theta}_p - \tilde{\theta}\right\| + \left\|\tilde{\theta} - \bar{\theta}\right\| = \left\|\tilde{\theta} - \bar{\theta}\right\| + \hat{k}.$$

We already have $\sqrt{n}\hat{k} = O_{\mathcal{Q}}(1)$, so it suffices to show $\sqrt{n}|\tilde{\theta} - \bar{\theta}| = O_{\mathcal{Q}}(1)$. Using Assumption D.4, we have

$$\begin{aligned}
C \min\left\{\delta, \|\tilde{\theta} - \bar{\theta}\|\right\} &\leq \max_{j=1,\dots,J} \left\{ \frac{E_P(m_j(X_i,\tilde{\theta}))}{\sigma_{P,j}(\tilde{\theta})} \right\} \\
&= \max_{j=1,\dots,J} \left\{ \frac{E_P(m_j(X_i,\hat{\theta}_p))}{\sigma_{P,j}(\hat{\theta}_p)} + \hat{k} D_{P_j}(\check{\theta}_j) p \right\} \\
&\leq \max_{j=1,\dots,J} \left\{ \frac{E_P(m_j(X_i,\hat{\theta}_p))}{\sigma_{P,j}(\hat{\theta}_p)} \right\} + \max_{j=1,\dots,J} \left\{ \hat{k} D_{P_j}(\check{\theta}_j) p \right\}.
\end{aligned}$$

Here, the equality step uses that $\tilde{\theta} = \hat{\theta}_p + \hat{k}p$ and introduces $\check{\theta}_j$, which lies componentwise between $\tilde{\theta}$ and $\hat{\theta}_p$. In the last line, the first term equals 0 w.p.a. 1 because $\hat{\theta}_p \in \hat{\Theta}_I$, and the second term is bounded by $\hat{k}\bar{M}$, hence the result. To justify Simplification 2, combine the above algebra with the following observations:

(i) For a sequence $P_n \in \mathcal{Q}$, coverage of $p'\theta$ for some $\theta \in H(p,\Theta_I(P_n))$ implies coverage of $s(p,\Theta_I(P_n))$. In the proof of Theorem 4.1, starting with display D.7, it therefore suffices to show the claim for some, possibly data dependent, sequence $\theta_n \in H(p,\Theta_I(P_n))$, and then again (in case of two-sided testing) for a sequence $\theta_n \in H(-p,\Theta_I(P_n))$.

(ii) All proofs go through if coverage is evaluated at θ_n but $D_{P,j}$ and $\mathbb{G}_{n,j}$ are estimated at some $\hat{\theta}_{n,p} = \theta_n + O_{\mathcal{Q}}(n^{-1/2})$. To give one example, Assumption 4.4 implies that $\|\hat{D}_{n,j}(\hat{\theta}_{n,p}) - D_{P_n,j}(\theta_n)\| = o_{\mathcal{Q}}(1)$.

Part III

This is established by showing that Assumption D.1-II is implied. Thus, let $\mathfrak{W}$ be as in (D.32). Because the marginals of $\mathbb{Z}$ are standard normal, for any $\eta > 0$ we have the Bonferroni bounds

$$\Pr\left(\mathfrak{W}(\bar{c}) \subseteq L_\eta\right) \geq 1 - \eta,$$

where

$$\begin{aligned} L_\eta &= \left\{\lambda \in \mathbb{R}^d : p'\lambda = 0 \cap \max_j \left\{\Phi^{-1}(\eta/J) + D_{P_j}\lambda\right\} \leq \bar{c}\right\} \\ &= \left\{\lambda \in \mathbb{R}^d : p'\lambda = 0 \cap \max_j D_{P_j}\lambda \leq \bar{c} + \Phi^{-1}(1-\eta/J)\right\}. \end{aligned}$$

It remains to bound $\|L_\eta\|_H = \max\{\|\lambda\| : \lambda \in L_\eta\}$. To do so, we show below that

$$p'\lambda = 0 \Rightarrow \max_j \left\{D_{P_j}\lambda / \|\lambda\|\right\} \geq \underbrace{\frac{\sqrt{1+\bar{\delta}^2}-1}{\sqrt{1+\bar{\delta}^2}+1}}_{=:a}\varepsilon_D, \tag{D.37}$$

where $\bar{\delta}$ is from Assumption D.5 and ε_D is from Assumption D.2. Solving (D.37) for $\|\lambda\|$ and inspecting the definition of L_η yields

$$\max\{\|\lambda\| : \lambda \in L_\eta\} \leq \frac{\bar{c} + \Phi^{-1}(1-\eta/J)}{a\varepsilon_D}$$

and therefore an $O(1)$ upper bound on $\|\mathfrak{W}(\bar{c})\|$. It remains to show (D.37). Suppose by contradiction that $\max_j \{D_j\lambda/\|\lambda\|\} < a\varepsilon_D$. Let the unit vector t^* achieve the minimum from Assumption D.2-2, then $\max_j \{D_j(\lambda/\|\lambda\| + dt^*)\} < 0$ and therefore $t \equiv \lambda/\|\lambda\| + dt^* \in \mathcal{T}$. We compute

$$\frac{\lambda' t}{\|\lambda\|\,\|t\|} = \frac{\lambda'\left(\frac{\lambda}{\|\lambda\|} + at^*\right)}{\|\lambda\|\left\|\frac{\lambda}{\|\lambda\|} + at^*\right\|} = \frac{1 + a\frac{\lambda' t^*}{\|\lambda\|}}{\left\|\frac{\lambda}{\|\lambda\|} + at^*\right\|} > \frac{1-a}{1+a} = 1/\sqrt{1+\bar{\delta}^2},$$

where the inequality is strict because $\lambda \neq t^*$. We conclude that $\max_{t\in\mathcal{T}} \frac{\lambda' t}{\|\lambda\|\|t\|} > 1/\sqrt{1+\bar{\delta}^2}$. In particular, if $\hat{\lambda}$ is the projection of λ onto $\mathcal{T}$, then $\frac{\lambda'\hat{\lambda}}{\|\lambda\|\|\hat{\lambda}\|} > 1/\sqrt{1+\bar{\delta}^2}$.[6]

However, we also have $p'\hat{\lambda}/\|\hat{\lambda}\| \leq -\bar{\delta}$ by Assumption D.5. It follows that $p'(\lambda - \hat{\lambda})/\|\hat{\lambda}\| \geq \bar{\delta}$, hence $\|\lambda - \hat{\lambda}\|^2 \geq \bar{\delta}^2\|\hat{\lambda}\|^2$ by Cauchy-Schwarz (recall p is a unit vector). But also $\|\lambda - \hat{\lambda}\|^2 + \|\hat{\lambda}\|^2 = \|\lambda\|^2$. Simple algebra then yields $\|\hat{\lambda}\|/\|\lambda\| \leq 1/\sqrt{1+\bar{\delta}^2}$. But $\|\hat{\lambda}\|/\|\lambda\|$ is also the cosine of the angle formed by λ and $\hat{\lambda}$. Thus, $\frac{\lambda'\hat{\lambda}}{\|\lambda\|\|\hat{\lambda}\|} \leq 1/\sqrt{1+\bar{\delta}^2}$, a contradiction.[7]

D.3.3 Proof of Theorem 4.4: Asymptotic Equivalence with BCS-Profiling in Well-Behaved Cases

Recall that under this Theorem's assumptions, $H(p, \Theta_I)$ is a singleton $\{\theta_p^*\}$ whose element is $\sqrt{n}$-consistently estimated by a sample analog $\hat{\theta}_p$. We restrict attention to $s \geq p'(\theta_p^* + t^*\varepsilon_D/2M)$, where terms are as in the proof of Theorem 4.3-(I). The proof for $s < p'(\theta_p^* + t^*\varepsilon_D/2M)$ is analogous. Similarly to earlier proofs, consider a sequence (P_n, s_n) that asymptotically minimizes the probability from the Theorem. If $\sqrt{n}(s_n - s(p, \Theta_I(P_n))) \to \infty$, then $\min_{p'\theta = s_n} T_n(\theta) \to \infty$ by arguments in the proof of Theorem 4.3-(II), and the conclusion obtains because both

[6] Verbally, if λ is near tangential to all constraints, it is near tangential to $\mathcal{T}$. The counterexample to this would be a "spike," which is excluded by Assumption D.2-2.

[7] Verbally, if $p'\lambda = 0$, then λ cannot be near tangential to $\mathcal{T}$ because of the "pointy cone" assumption D.5, yielding a contradiction.

indicator functions vanish. Similarly, if $\sqrt{n}(s_n - s(p,\Theta_I(P_n))) \to -\infty$, then both indicator functions equal 1 with probability approaching 1 (indeed, recall the sample support function is $\sqrt{n}$-consistent). It remains to analyze the case where $\sqrt{n}(s_n - s(p,\Theta_I(P_n))) = O_{\mathcal{Q}}(1)$.

Recalling that no ρ-box is used, $\hat{c}_n(\hat{\theta}_p)$ is the $(1-\alpha)$ quantile of

$$\begin{aligned} T_n^b &= \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\hat{\theta}_p) + \kappa_n^{-1}\sqrt{n}\big|\bar{m}_{n,j}(\hat{\theta}_p)/\hat{\sigma}_{n,j}(\hat{\theta}_p)\big|_- + \hat{D}_{n,j}(\hat{\theta}_p)\lambda\right\} \\ &\overset{(1)}{=} \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\hat{\theta}_p) + \kappa_n^{-1}\sqrt{n}E_P\big|m_j(X_i,\hat{\theta}_p)/\sigma_{P,j}(\hat{\theta}_p)\big|_- + \hat{D}_{n,j}(\hat{\theta}_p)\lambda\right\} + o_{\mathcal{Q}}(1) \\ &\overset{(2)}{=} \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\theta_p^*) + \kappa_n^{-1}\sqrt{n}E_P\big|m_j(X_i,\theta_p^*)/\sigma_{P,j}(\theta_p^*)\big|_- + D_{n,j}(\theta_p^*)\lambda\right\} + o_{\mathcal{Q}}(1), \end{aligned}$$

Here, (1) uses Lemma E.5-(iii). Step (2) uses that by Theorem 4.3-(III), the values of λ solving the optimization problems are $O_{\mathcal{Q}}(1)$; by 4.3-(II), $\sqrt{n}(\hat{\theta}_p - \theta_p^*) = O_{\mathcal{Q}}(1)$; and smoothness conditions as well as consistent estimation of gradients. These jointly imply that $|\hat{D}_{n,j}(\hat{\theta}_p)\lambda - D_{n,j}(\theta_p^*)\lambda| = o_{\mathcal{Q}}(1)$ uniformly over the relevant range of λ.

To compare BCS-profiling, let $\hat{\theta}_{p,s_n}$ be the selection from $\arg\min_{p'\theta=s_n}|\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_+$ that solves the problem in the definition of $T_n^{DR}(s_n)$ below. Arguments very similar to Theorem 4.3-(II) imply that $\sqrt{n}(\hat{\theta}_{p,s_n} - \theta_p^*) = O_{\mathcal{Q}}(1)$. We can use this, again Lemma E.5-(iii), and smoothness conditions to write

$$\begin{aligned} T_n^{DR}(s_n) &= \min_\theta\max_j \left\{\mathbb{G}_{n,j}^b(\theta) + \kappa_n^{-1}\sqrt{n}\,|\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_-\right\} \quad s.t.\ \ \theta \in \arg\min_{p'\theta=s_n}\max_j |\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|_+ \\ &= \max_j \left\{\mathbb{G}_{n,j}^b(\hat{\theta}_{p,s_n}) + \kappa_n^{-1}\sqrt{n}\big|\bar{m}_{n,j}(\hat{\theta}_{p,s_n})/\hat{\sigma}_{n,j}(\hat{\theta}_{p,s_n})\big|_-\right\} \\ &= \max_j \left\{\mathbb{G}_{n,j}^b(\hat{\theta}_{p,s_n}) + \kappa_n^{-1}\sqrt{n}E_P\big|m_j(X_i,\hat{\theta}_{p,s_n})/\sigma_{P,j}(\hat{\theta}_{p,s_n})\big|_-\right\} + o_{\mathcal{Q}}(1) \\ &= \max_j \left\{\mathbb{G}_{n,j}^b(\theta_p^*) + \kappa_n^{-1}\sqrt{n}E_P\left|m_j(X_i,\theta_p^*)/\sigma_{P,j}(\theta_p^*)\right|_-\right\} + o_{\mathcal{Q}}(1). \end{aligned}$$

Next,

$$\begin{aligned} T_n^{PR}(s_n) &= \min_{\theta\in\Theta:p'\theta=s_n}\max_j \left\{\mathbb{G}_{n,j}^b(\theta) + \kappa_n^{-1}\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)\right\} \\ &\overset{(1)}{=} \min_{\theta\in\Theta:p'\theta=s_n}\max_j \left\{\mathbb{G}_{n,j}^b(\theta) + \kappa_n^{-1}\sqrt{n}E_P\left(m_j(X_i,\theta)/\sigma_{P,j}(\theta)\right)\right\} + o_{\mathcal{Q}}(1) \\ &\overset{(2)}{=} \min_{\theta\in\Theta:p'\theta=s(p,\Theta_I)}\max_j \left\{\mathbb{G}_{n,j}^b(\theta) + \kappa_n^{-1}\sqrt{n}E_P\left(m_j(X_i,\theta)/\sigma_{P,j}(\theta)\right)\right\} + o_{\mathcal{Q}}(1) \\ &\overset{(3)}{=} \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\theta_p^* + \lambda\kappa_n n^{-1/2}) + \kappa_n^{-1}\sqrt{n}E_P\big(m_j(X_i,\theta_p^* + \lambda\kappa_n n^{-1/2})/\sigma_{P,j}(\theta_p^* + \lambda\kappa_n n^{-1/2})\big)\right\} + o_{\mathcal{Q}}(1) \\ &\overset{(4)}{=} \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\theta_p^*) + \kappa_n^{-1}\sqrt{n}E_P\left(m_j(X_i,\theta_p^*)/\sigma_{P,j}(\theta_p^*)\right) + D_{P,j}(\theta_p^*)\lambda\right\} + o_{\mathcal{Q}}(1) \\ &\overset{(5)}{=} \min_{p'\lambda=0}\max_j \left\{\mathbb{G}_{n,j}^b(\theta_p^*) + \kappa_n^{-1}\sqrt{n}E_P\left|m_j(X_i,\theta_p^*)/\sigma_{P,j}(\theta_p^*)\right|_- + D_{P,j}(\theta_p^*)\lambda\right\} + o_{\mathcal{Q}}(1) \end{aligned}$$

Here, (1) uses Lemma E.5-(iii). The first crucial step is (2), which uses that the distance between the hyperplanes $\{p'\theta = s_n\}$ and $\{p'\theta = s(\Theta_I,p)\}$ is of order $O_{\mathcal{Q}}(n^{-1/2})$, together with smoothness conditions. Step (3) reparameterizes $\theta = \theta_p^* + \lambda\kappa_n n^{-1/2}$. Crucially, BCS prove that the λ solving the problem is $O_{\mathcal{Q}}(1)$. This means the problem can be uniformly linearized, justifying step (4). Step (4) also observes cancellation of factors multiplying $D_{P,j}(\theta_p^*)\lambda$. Step (5) uses that $\theta_p^* \in \Theta_I$. Finally, Assumption 4.3 ensures that the true distribution of T_n, as well as the above approximations, are of order $O_{\mathcal{Q}}(1)$. We conclude that $T_n^{PR}(s_n)$ asymptotically agrees with, and $T_n^{DR}(s_n)$ asymptotically dominates, T_n^b. □

Appendix E Auxiliary Lemmas

E.1 Lemmas Used to Prove Theorems 4.1 and 4.2

Throughout this Appendix, we let $(P_n, \theta_n) \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ be a subsequence as defined in the proof of Theorem 4.1. That is, along (P_n, θ_n), one has

$$\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]},\ j = 1, \dots, J, \tag{E.1}$$

$$\Omega_{P_n} \overset{u}{\to} \Omega, \tag{E.2}$$

$$D_{P_n}(\theta_n) \to D. \tag{E.3}$$

Fix $c \geq 0$. For each $\lambda \in \mathbb{R}^d$ and $\theta \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$, let

$$\mathfrak{w}_j(\lambda) \equiv \mathbb{Z}_j + \rho D_j \lambda + \pi^*_{1,j}, \tag{E.4}$$

where $\pi^*_{1,j}$ is defined in (D.5) and we used Lemma E.5. Under Assumption 4.3-(II) if

$$\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1}, \tag{E.5}$$

we replace the constraints

$$\mathbb{Z}_j + \rho D_j \lambda \leq c, \tag{E.6}$$

$$\mathbb{Z}_{j+R_1} + \rho D_{j+R_1} \lambda \leq c, \tag{E.7}$$

with

$$\mu_j(\theta)\{\mathbb{Z}_j + \rho D_j \lambda\} - \mu_{j+R_1}(\theta)\{\mathbb{Z}_{j+R_1} + \rho D_{j+R_1}\lambda\} \leq c, \tag{E.8}$$

$$-\mu_j(\theta)\{\mathbb{Z}_j + \rho D_j \lambda\} + \mu_{j+R_1}(\theta)\{\mathbb{Z}_{j+R_1} + \rho D_{j+R_1}\lambda\} \leq c, \tag{E.9}$$

where

$$\mu_j(\theta) = \begin{cases} 1 & \text{if } \gamma_{1,P_n,j}(\theta) = 0 = \gamma_{1,P_n,j+R_1}(\theta), \\ \frac{\gamma_{1,P_n,j+R_1}(\theta)}{\gamma_{1,P_n,j+R_1}(\theta) + \gamma_{1,P_n,j}(\theta)} & \text{otherwise}, \end{cases} \tag{E.10}$$

$$\mu_{j+R_1}(\theta) = \begin{cases} 0 & \text{if } \gamma_{1,P_n,j}(\theta) = 0 = \gamma_{1,P_n,j+R_1}(\theta), \\ \frac{\gamma_{1,P_n,j}(\theta)}{\gamma_{1,P_n,j+R_1}(\theta) + \gamma_{1,P_n,j}(\theta)} & \text{otherwise}, \end{cases} \tag{E.11}$$

When Assumption 4.3-(II) is invoked with hard-threshold GMS, replace constraints j and $j+R_1$ in the definition of $\Lambda_n^b(\theta'_n, \rho, c), \theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ in equation (3.1) as described on p.14 of the paper; when it is invoked with a GMS function φ that is smooth in its argument, replace them, respectively, with

$$\hat{\mu}_{n,j}(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \hat{D}_{n,j}(\theta'_n)\lambda\} - \hat{\mu}_{n,j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \hat{D}_{n,j+R_1}(\theta'_n)\lambda\} + \varphi_j(\hat{\xi}_{n,j}(\theta'_n)) \leq c, \tag{E.12}$$

$$-\hat{\mu}_{n,j}(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \hat{D}_{n,j}(\theta'_n)\lambda\} + \hat{\mu}_{n,j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \hat{D}_{n,j+R_1}(\theta'_n)\lambda\} + \varphi_{j+R_1}(\hat{\xi}_{n,j+R_1}(\theta'_n)) \leq c, \tag{E.13}$$

where

$$\hat{\mu}_{n,j+R_1}(\theta'_n) = \min\left\{\max\left(0, \frac{\frac{\bar{m}_{n,j}(\theta'_n)}{\hat{\sigma}_{n,j}(\theta'_n)}}{\frac{\bar{m}_{n,j+R_1}(\theta'_n)}{\hat{\sigma}_{n,j+R_1}(\theta'_n)} + \frac{\bar{m}_{n,j}(\theta'_n)}{\hat{\sigma}_{n,j}(\theta'_n)}}\right), 1\right\}, \tag{E.14}$$

$$\hat{\mu}_{n,j}(\theta'_n) = 1 - \hat{\mu}_{n,j+R_1}(\theta'_n). \tag{E.15}$$

Let $\mathfrak{B}^d_\rho = \lim_{n\to\infty} B^d_{n,\rho}$. Let the intersection of $\{\lambda \in \mathfrak{B}^d_\rho : \ p'\lambda = 0\}$ with the level set associated with the so defined function $\mathfrak{w}_j(\lambda)$ be

$$\mathfrak{W}(c) \equiv \left\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c, \ \forall j = 1,\dots,J\right\}. \tag{E.16}$$

Due to the substitutions in equations (E.6)-(E.9), the paired inequalities (i.e., inequalities for which (E.5) holds under Assumption 4.3-(II)) are now genuine equalities relaxed by c. With some abuse of notation, we index them among the $j = J_1+1,\dots,J$. With that convention, for given $\delta \in \mathbb{R}$, define

$$\mathfrak{W}^\delta(c) \equiv \left\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c+\delta, \ \forall j = 1,\dots,J_1, \right. \\ \left. \cap\, \mathfrak{w}_j(\lambda) \le c, \ \forall j = J_1+1,\dots,J\right\}. \tag{E.17}$$

Define the $(J+2d+2)\times d$ matrix

$$K_P(\theta,\rho) \equiv \begin{bmatrix} [\rho D_{P,j}(\theta)]_{j=1}^{J_1+J_2} \\ [-\rho D_{P,j-J_2}(\theta)]_{j=J_1+J_2+1}^{J} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}. \tag{E.18}$$

Given a square matrix A, we let $\mathrm{eig}(A)$ denote its smallest eigenvalue. In all Lemmas below, we assume $\alpha < 1/2$.

Lemma E.1: *Let Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. Let $\{P_n,\theta_n\}$ be a sequence such that $P_n \in \mathcal{P}$ and $\theta_n \in \Theta_I(P_n)$ for all n and $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}$, $j = 1,\dots,J$, $\Omega_{P_n} \overset{u}{\to} \Omega$, and $D_{P_n}(\theta_n) \to D$. Then,*

$$\liminf_{n\to\infty} P_n\left(U_n(\theta_n,\hat{c}_{n,\rho}) \neq \emptyset\right) \ge 1-\alpha. \tag{E.19}$$

Proof. We consider a subsequence along which $\liminf_{n\to\infty} P_n(U_n(\theta_n,\hat{c}_{n,\rho} \neq \emptyset)$ is achieved as a limit. For notational simplicity, we use $\{n\}$ for this subsequence below.

Below, we construct a sequence of critical values such that

$$\hat{c}_n(\theta'_n) \ge c^I_n(\theta'_n) + o_{P_n}(1), \tag{E.20}$$

and $c^I_n(\theta'_n) \overset{P_n}{\to} c_{\pi^*}$ for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$. The construction is as follows. When $c_{\pi^*} = 0$, let $c^I_n(\theta'_n) = 0$ for all $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$, and hence $c^I_n(\theta'_n) \overset{P_n}{\to} c_{\pi^*}$. If $c_{\pi^*} > 0$, let $c^I_n(\theta_n) \equiv \inf\{c \in \mathbb{R}_+ : P^*_n(V^I_n(\theta_n,c)) \ge 1-\alpha\}$, where V^I_n is defined as in Lemma E.3. By Lemma E.3 (iii), this critical value sequence satisfies (E.20) with probability approaching 1. Further, by Lemma E.3 (ii), $c^I_n(\theta'_n) \overset{P_n}{\to} c_{\pi^*}$ for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$.

For each $\theta \in \Theta$, let

$$c^I_{n,\rho}(\theta) \equiv \inf_{\lambda \in B^d_{n,\rho}} c^I_n(\theta + \frac{\lambda\rho}{\sqrt{n}}). \tag{E.21}$$

Since the $o_{P_n}(1)$ term in (E.20) does not affect the argument below, we redefine $c^I_{n,\rho}(\theta_n)$ as $c^I_{n,\rho}(\theta_n) + o_{P_n}(1)$. By

(E.20) and simple addition and subtraction,

$$P_n\Big(U_n(\theta_n,\hat{c}_{n,\rho}(\theta_n))\neq\emptyset\Big)\geq P_n\Big(U_n(\theta_n,c^I_{n,\rho}(\theta_n))\neq\emptyset\Big)$$
$$=\Pr(\mathfrak{W}(c_{\pi^*})\neq\emptyset)+\Big[P_n\Big(U_n(\theta_n,c^I_{n,\rho}(\theta_n))\neq\emptyset\Big)-\Pr\Big(\mathfrak{W}(c_{\pi^*})\neq\emptyset\Big)\Big]. \tag{E.22}$$

As previously argued, $\mathbb{G}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}})\xrightarrow{d}\mathbb{Z}$. Moreover, by Lemma E.10, $\sup_{\theta\in\Theta}\|\eta_n(\theta)\|\xrightarrow{p}0$ uniformly in $\mathcal{P}$, and by Lemma E.3, $c^I_{n,\rho}(\theta_n)\xrightarrow{p}c_{\pi^*}$. Therefore, uniformly in $\lambda\in B^d$, the sequence $\{(\mathbb{G}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\eta_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),c^I_{n,\rho}(\theta_n))\}$ satisfies

$$(\mathbb{G}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\eta_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),c^I_{n,\rho}(\theta_n))\xrightarrow{d}(\mathbb{Z},0,c_{\pi^*}). \tag{E.23}$$

In what follows, using Lemma 1.10.4 in van der Vaart and Wellner (2000) we take $(\mathbb{G}^*_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\eta^*_n,c^*_n)$ to be the almost sure representation of $(\mathbb{G}_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\eta_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),c^I_{n,\rho}(\theta_n))$ defined on some probability space $(\Omega,\mathcal{F},\mathbf{P})$ such that $(\mathbb{G}^*_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\eta^*_n,c^*_n)\xrightarrow{a.s.}(\mathbb{Z}^*,0,c_{\pi^*})$, where $\mathbb{Z}^*\overset{d}{=}\mathbb{Z}$.

For each $\lambda\in\mathbb{R}^d$, we define analogs to the quantities in (D.24) and (E.4) as

$$u^*_{n,j,\theta_n}(\lambda)\equiv\{\mathbb{G}^*_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar{\theta}_n)\lambda+\pi^*_{1,j}\}(1+\eta^*_{n,j}), \tag{E.24}$$
$$\mathfrak{w}^*_j(\lambda)\equiv\mathbb{Z}^*_j+\rho D_j\lambda+\pi^*_{1,j}. \tag{E.25}$$

where we used that by Lemma E.5, $\kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\theta_n)-\kappa_n^{-1}\sqrt{n}\gamma_{1,P,j}(\theta'_n)=o(1)$ uniformly over $\theta'_n\in(\theta_n+\rho/\sqrt{n}B^d)\cap\Theta$ and therefore $\pi^*_{1,j}$ is constant over this neighborhood, and we applied a similar replacement as described in equations (E.6)-(E.9) for the case that $\pi^*_{1,j}=0=\pi^*_{1,j+R_1}$. Similarly, we define analogs to the sets in (D.25) and (E.16) as

$$U^*_n(\theta_n,c^*_n)\equiv\big\{\lambda\in B^d_{n,\rho}:p'\lambda=0\cap u^*_{n,j,\theta_n}(\lambda)\leq c^*_n,\ \forall j=1,\dots,J\big\}, \tag{E.26}$$
$$\mathfrak{W}^*(c_{\pi^*})\equiv\big\{\lambda\in\mathfrak{B}^d_\rho:p'\lambda=0\cap\mathfrak{w}^*_j(\lambda)\leq c_{\pi^*},\ \forall j=1,\dots,J\big\}. \tag{E.27}$$

It then follows that equation (E.22) can be rewritten as

$$P_n\Big(U_n(\theta_n,\hat{c}_{n,\rho}(\theta_n))\neq\emptyset\Big)\geq\mathbf{P}(\mathfrak{W}^*(c_{\pi^*})\neq\emptyset)+\Big[\mathbf{P}\Big(U^*_n(\theta_n,c^*_n)\neq\emptyset\Big)-\mathbf{P}\Big(\mathfrak{W}^*(c_{\pi^*})\neq\emptyset\Big)\Big]. \tag{E.28}$$

By the definition of c_{π^*}, we have $\mathbf{P}(\mathfrak{W}^*(c_{\pi^*})\neq\emptyset)\geq1-\alpha$. Therefore, we are left to show that the second term on the right hand side of (E.28) tends to 0 as $n\to\infty$.

Define

$$\mathcal{J}^*\equiv\{j=1,\cdots,J:\pi^*_{1,j}=0\}. \tag{E.29}$$

Case 1. Suppose first that $\mathcal{J}^*=\emptyset$, which implies $J_2=0$ and $\pi^*_{1,j}=-\infty$ for all j. Then we have

$$U^*_n(\theta_n,c^*_n)=\{\lambda\in B^d_{n,\rho}:p'\lambda=0\},\quad\mathfrak{W}^*(c_{\pi^*})=\{\lambda\in\mathfrak{B}^d_\rho:p'\lambda=0\}, \tag{E.30}$$

with probability 1, and hence

$$\mathbf{P}\Big(\{U^*_n(\theta_n,c^*_n)\neq\emptyset\}\cap\{\mathfrak{W}^*(c_{\pi^*})\neq\emptyset\}\Big)=1. \tag{E.31}$$

This in turn implies that

$$\Big|\mathbf{P}\Big(U^*_n(\theta_n,c^*_n)\neq\emptyset\}\Big)-\mathbf{P}\Big(\mathfrak{W}^*(c_{\pi^*})\neq\emptyset\}\Big)\Big|=0, \tag{E.32}$$

where we used $|\mathbf{P}(A)-\mathbf{P}(B)|\le\mathbf{P}(A\Delta B)\le 1-\mathbf{P}(A\cap B)$ for any pair of events A and B. Hence, the term in the square brackets in (E.28) is 0.

Case 2. Now consider the case that $\mathcal{J}^*\neq\emptyset$. We show that the term in the square brackets in (E.28) converges to 0. To that end, note that for any events A,B,

$$\Big|\mathbf{P}(A\neq\emptyset)-\mathbf{P}(B\neq\emptyset)\Big|\le\Big|\mathbf{P}(\{A=\emptyset\}\cap\{B\neq\emptyset\})+\mathbf{P}(\{A\neq\emptyset\}\cap\{B=\emptyset\})\Big| \tag{E.33}$$

Hence, we aim to establish that for $A=U_n^*(\theta_n,c_n^*)$, $B=\mathfrak{W}^*(c_{\pi^*})$, the right hand side of equation (E.33) converges to zero. But this is guaranteed by Lemma E.2. Therefore, the conclusion of the lemma follows. □

Lemma E.2: *Let Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. Let (P_n,θ_n) have the almost sure representations given in Lemma E.1, and let $\mathcal{J}^*$ be defined as in (E.29). Assume that $\mathcal{J}^*\neq\emptyset$. Then for any $\eta>0$, there exists $N\in\mathbb{N}$ such that*

$$\mathbf{P}\Big(\{U_n^*(\theta_n,c_n^*)\neq\emptyset\}\cap\{\mathfrak{W}^*(c_{\pi^*})=\emptyset\}\Big)\le\eta/2, \tag{E.34}$$

$$\mathbf{P}\Big(\{U_n^*(\theta_n,c_n^*)=\emptyset\}\cap\{\mathfrak{W}^*(c_{\pi^*})\neq\emptyset\}\Big)\le\eta/2, \tag{E.35}$$

for all $n\ge N$, where the sets in the above expressions are defined in equations (E.26) and (E.27).

Proof. We begin by observing that for $j\notin\mathcal{J}^*$, $\pi_{1,j}^*=-\infty$, and therefore the corresponding inequalities

$$\left(\mathbb{G}_{n,j}^*(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar{\theta}_n)\lambda+\pi_{1,j}^*\right)(1+\eta_{n,j}^*)\le c_n^*,$$
$$\mathbb{Z}_j^*+\rho D_j\lambda+\pi_{1,j}^*\le c_{\pi^*}$$

are satisfied with probability approaching one by similar arguments as in (D.20). Hence, we can redefine the sets of interest as

$$U_n^*(\theta_n,c_n^*)\equiv\big\{\lambda\in B_{n,\rho}^d:p'\lambda=0\cap u_{n,j,\theta_n}^*(\lambda)\le c_n^*,\ \forall j\in\mathcal{J}^*\big\}, \tag{E.36}$$

$$\mathfrak{W}^*(c_{\pi^*})\equiv\big\{\lambda\in\mathfrak{B}_\rho^d:p'\lambda=0\cap\mathfrak{w}_j^*(\lambda)\le c_{\pi^*},\ \forall j\in\mathcal{J}^*\big\}. \tag{E.37}$$

We first show (E.34). For this, we start by defining the events

$$A_n\equiv\left\{\sup_{\lambda\in B^d}\max_{j\in\mathcal{J}^*}\big|(u_{n,j,\theta_n}^*(\lambda)-c_n^*)-(\mathfrak{w}_j^*(\lambda)-c_{\pi^*})\big|\ge\delta\right\}. \tag{E.38}$$

By Lemma E.4, using the assumption that $\mathcal{J}^*\neq\emptyset$, for any $\eta>0$ there exists $N\in\mathbb{N}$ such that

$$\mathbf{P}(A_n)<\eta/2,\ \forall n\ge N. \tag{E.39}$$

Define the sets of λs, $U_n^{*,+\delta}$ and $\mathfrak{W}^{*,+\delta}$ by relaxing the constraints shaping U_n^* and $\mathfrak{W}^*$ by δ:

$$U_n^{*,+\delta}(\theta_n,c)\equiv\{\lambda\in B_{n,\rho}^d:p'\lambda=0\cap u_{n,j,\theta_n}^*(\lambda)\le c+\delta,\ j\in\mathcal{J}^*\}, \tag{E.40}$$

$$\mathfrak{W}^{*,+\delta}(c)\equiv\{\lambda\in\mathfrak{B}_\rho^d:p'\lambda=0\cap\mathfrak{w}_j^*(\lambda)\le c+\delta,\ j\in\mathcal{J}^*\}. \tag{E.41}$$

Compared to the set in equation (E.17), here we replace $u_{n,j,\theta_n}^*(\lambda)$ for $u_{n,j,\theta_n}(\lambda)$ and $\mathfrak{w}_j^*(\lambda)$ for $\mathfrak{w}_j(\lambda)$, we retain only constraints in $\mathcal{J}^*$, and we relax all such constraints by $\delta>0$ instead of relaxing only those in $\{1,\dots,J_1\}$. Next, define the event $L_n\equiv\{U_n^*(\theta_n,c_n^*)\subset\mathfrak{W}^{*,+\delta}(c_{\pi^*})\}$ and note that $A_n^c\subseteq L_n$.

We may then bound the left hand side of (E.34) as

$$\mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big) \leq \mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) \neq \emptyset\} \cap \{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) = \emptyset\}\Big) + \mathbf{P}\Big(\{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big), \tag{E.42}$$

where we used $P(A \cap B) \leq P(A \cap C) + P(B \cap C^c)$ for any events A, B, and C. The first term on the right hand side of (E.42) can further be bounded as

$$\mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) \neq \emptyset\} \cap \{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) = \emptyset\}\Big) \leq \mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) \not\subseteq \mathfrak{W}^{*,+\delta}(c_{\pi^*})\}\Big) = \mathbf{P}(L_n^c) \leq \mathbf{P}(A_n) < \eta/2, \ \forall n \geq N \ , \tag{E.43}$$

where the penultimate inequality follows from $A_n^c \subseteq L_n$ as argued above, and the last inequality follows from (E.39). For the second term on the left hand side of (E.42), by Lemma E.6, there exists $N' \in \mathbb{N}$ such that

$$\mathbf{P}\Big(\{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big) \leq \eta/2, \ \forall n \geq N'. \tag{E.44}$$

Hence, (E.34) follows from (E.42), (E.43), and (E.44).

To establish (E.35), we distinguish three cases.

Case 1. Suppose first that $J_2 = 0$ (recalling that under Assumption 4.3-(II) this means that there is no $j = 1, \dots, R_1$ such that $\pi_{1,j}^* = 0 = \pi_{1,j+R_1}^*$), and hence one has only moment inequalities. In this case, by (E.36) and (E.37), one may write

$$U_n^*(\theta_n, c) \equiv \left\{\lambda \in B_{n,\rho}^d : p'\lambda = 0 \cap u_{n,j,\theta_n}^*(\lambda) \leq c, \ j \in \mathcal{J}^*\right\}, \tag{E.45}$$

$$\mathfrak{W}^{*,-\delta}(c) \equiv \left\{\lambda \in \mathfrak{B}_\rho^d : p'\lambda = 0 \cap \mathfrak{w}_j^*(\lambda) \leq c - \delta, \ j \in \mathcal{J}^*\right\}, \tag{E.46}$$

where $\mathfrak{W}^{*,-\delta}$, $\delta > 0$, is obtained by tightening the inequality constraints shaping $\mathfrak{W}^*$. Define the event

$$R_{2n} \equiv \{\mathfrak{W}^{*,-\delta}(c_{\pi^*}) \subset U_n^*(\theta_n, c_n^*)\}, \tag{E.47}$$

and note that $A_n^c \subseteq R_{2n}$. The result in equation (E.35) then follows by Lemma E.6 using again similar steps to (E.42)-(E.44).

Case 2. Next suppose that $J_2 \geq d$. In this case, we define $\mathfrak{W}^{*,-\delta}$ to be the set obtained by tightening by δ the inequality constraints as well as each of the two opposing inequalities obtained from the equality constraints. That is,

$$\mathfrak{W}^{*,-\delta}(c_{\pi^*}) \equiv \{\lambda \in \mathfrak{B}_\rho^d : p'\lambda = 0 \cap \mathfrak{w}_j^*(\lambda) \leq c - \delta, \ j \in \mathcal{J}^*\}, \tag{E.48}$$

that is, the same set as in (E.133) with $\mathfrak{w}_j^*(\lambda)$ replacing $\mathfrak{w}_j(\lambda)$ and defining the set using only inequalities in $\mathcal{J}^*$. Note that, by Lemma E.8, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ $c_n^I(\theta)$ is bounded from below by some $\underline{c} > 0$ with probability approaching one uniformly in $P \in \mathcal{P}$ and $\theta \in \Theta_I(P)$. This ensures c_{π^*} is bounded from below by $\underline{c} > 0$. This in turn allows us to construct a non-empty tightened constraint set with probability approaching 1. Namely, for $\delta < \underline{c}$, $\mathfrak{W}^{*,-\delta}(c_{\pi^*})$ is nonempty with probability approaching 1 by Lemma E.6, and hence its superset $\mathfrak{W}^*(c_{\pi^*})$ is also non-empty with probability approaching 1. However, note that $A_n^c \subseteq R_{2n}$, where R_{2n} is in (E.47) now defined using the tightened constraint set $\mathfrak{W}^{*,-\delta}(c_{\pi^*})$ being defined as in (E.48), and therefore the same argument as in the previous case applies.

Case 3. Finally, suppose that $1 \leq J_2 < d$. Recall that, with probability 1 (under $\mathbf{P}$),

$$c_{\pi^*} = \lim_{n\to\infty} c_n^*, \tag{E.49}$$

and note that by construction $c_{\pi^*} \geq 0$. Consider first the case that $c_{\pi^*} > 0$. Then, by taking $\delta < c_{\pi^*}$, the argument in Case 2 applies.

Next consider the case that $c_{\pi^*} = 0$. Observe that

$$\mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) = \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) \neq \emptyset\}\Big) \leq \mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) = \emptyset\} \cap \{\mathfrak{W}^{*,-\delta}(0) \neq \emptyset\}\Big) \\ + \mathbf{P}\Big(\{\mathfrak{W}^{*,-\delta}(0) = \emptyset\} \cap \{\mathfrak{W}^*(0) \neq \emptyset\}\Big), \tag{E.50}$$

with $\mathfrak{W}^{*,-\delta}(0)$ defined as in (E.17) with $c = 0$ and with $\mathfrak{w}_j^*(\lambda)$ replacing $\mathfrak{w}_j(\lambda)$. By Lemma E.6, for any $\eta > 0$ there exists $\delta > 0$ and $N \in \mathbb{N}$ such that

$$\mathbf{P}\Big(\{\mathfrak{W}^{*,-\delta}(0) = \emptyset\} \cap \{\mathfrak{W}^*(0) \neq \emptyset\}\Big) < \eta/3 \ \forall n \geq N. \tag{E.51}$$

Therefore, the second term on the right hand side of (E.50) can be made arbitrarily small.

We now consider the first term on the right hand side of (E.50). Let g be a $J + 2d + 2$ vector with

$$g_j = \begin{cases} -\mathbb{Z}_j, & j \in \mathcal{J}^*, \\ 0, & j \in \{1, \cdots, J\} \setminus \mathcal{J}^*, \\ 1, & j = J+1, \ldots, J+2d, \\ 0, & j = J+2d+1, J+2d+2, \end{cases} \tag{E.52}$$

where we used that $\pi_{1,j}^* = 0$ for $j \in \mathcal{J}^*$ and where the last assignment is without loss of generality because of the considerations leading to the sets in (E.36)-(E.37).

For a given set $C \subset \{1, \ldots, J+2d+2\}$, let the vector g^C collect the entries of g^C corresponding to indices in C. Let

$$K \equiv \begin{bmatrix} [\rho D_j]_{j=1}^{J_1+J_2} \\ [-\rho D_{j-J_2}]_{j=J_1+J_2+1}^{J} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}. \tag{E.53}$$

Let the matrix K^C collect the rows of K corresponding to indices in C.

Let $\widetilde{\mathcal{C}}$ collect all size d subsets C of $\{1, ..., J+2d+2\}$ ordered lexicographically by their smallest, then second smallest, etc. elements. Let the random variable $\mathcal{C}$ equal the first element of $\widetilde{\mathcal{C}}$ s.t. $\det K^C \neq 0$ and $\lambda^C = (K^C)^{-1} g^C \in \mathfrak{W}^{*,-\delta}(0)$ if such an element exists; else, let $\mathcal{C} = \{J+1, ..., J+d\}$ and $\lambda^C = \mathbf{1}_d$, where $\mathbf{1}_d$ denotes a d vector with each entry equal to 1. Recall that $\mathfrak{W}^{*,-\delta}(0)$ is a (possibly empty) measurable random polyhedron in a compact subset of $\mathbb{R}^d$, see, e.g., Molchanov (2005, Definition 1.1.1). Thus, if $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$, then $\mathfrak{W}^{*,-\delta}(0)$ has extreme points, each of which is characterized as the intersection of d (not necessarily unique) linearly independent constraints interpreted as equalities. Therefore, $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$ implies that $\lambda^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)$ and therefore also that $\mathcal{C} \subset \mathcal{J}^* \cup \{J+1, \ldots, J+2d+2\}$. Note that the associated random vector $\lambda^{\mathcal{C}}$ is a measurable selection of a random closed set that equals $\mathfrak{W}^{*,-\delta}(0)$ if $\mathfrak{W}^{*,-\delta}(0) \neq \emptyset$ and equals $\mathfrak{B}_\rho^d$ otherwise, see, e.g., Molchanov (2005, Definition 1.2.2).

Lemma E.7 establishes that for any $\eta > 0$, there exist $\varepsilon_\eta > 0$ and N s.t. $n \geq N$ implies

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \eta, \tag{E.54}$$

which in turn, given our definition of $\mathcal{C}$, yields that there is $M > 0$ and N such that

$$\mathbf{P}\Big(\left|\det \left(K^{\mathcal{C}}\right)^{-1}\right| \leq M\Big) \geq 1 - \eta, \ \forall n \geq N. \tag{E.55}$$

Let g_n be a $J + 2d + 2$ vector with

$$g_{n,j}(\theta + \lambda/\sqrt{n}) \equiv \begin{cases} c_n^*/(1+\eta_{n,j}^*) - \mathbb{G}_{n,j}^*(\theta + \frac{\lambda\rho}{\sqrt{n}}) & \text{if } j \in \mathcal{J}^*, \\ 0, & \text{if } j \in \{1,\cdots,J\} \setminus \mathcal{J}^*, \\ 1, & \text{if } j = J+1,\ldots,J+2d, \\ 0, & \text{if } j = J+2d+1, J+2d+2, \end{cases} \tag{E.56}$$

using again that $\pi_{1,j}^* = 0$ for $j \in \mathcal{J}^*$. For each $P \in \mathcal{P}$, let

$$K_P(\theta,\rho) \equiv \begin{bmatrix} [\rho D_{P,j}(\theta)]_{j=1}^{J_1+J_2} \\ [-\rho D_{P,j-J_2}(\theta)]_{j=J_1+J_2+1}^{J} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}. \tag{E.57}$$

For each n and $\lambda \in B^d$, define the mapping $\phi_n : B^d \to \mathbb{R}^d_{[\pm\infty]}$ by

$$\phi_n(\lambda) \equiv \left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda),\rho)\right)^{-1} g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \tag{E.58}$$

where the notation $\bar\theta(\theta_n,\lambda)$ emphasizes that $\bar\theta$ depends on θ_n and λ because it lies component-wise between θ_n and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$. We show that ϕ_n is a contraction mapping and hence has a fixed point.

For any $\lambda, \lambda' \in B^d$ write

$$\begin{aligned} \|\phi_n(\lambda) - \phi_n(\lambda')\| &= \left\|\left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda),\rho)\right)^{-1} g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda'),\rho)\right)^{-1} g_n^{\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}})\right\| \\ &\leq \left\|\left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda),\rho)\right)^{-1}\right\|_2 \left\|g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - g_n^{\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}})\right\| \\ &\quad + \left\|\left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda),\rho)\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}(\bar\theta(\theta_n,\lambda'),\rho)\right)^{-1}\right\|_2 \left\|g_n^{\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}})\right\|, \end{aligned} \tag{E.59}$$

where $\|\cdot\|_2$ denotes the spectral norm (induced by the Euclidean norm).

By Assumption 4.5 (ii), for any $\eta > 0$, $k > 0$, there is $N \in \mathbb{N}$ such that

$$\begin{aligned} &\mathbf{P}\left(\left\|g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - g_n^{\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}})\right\| \leq k\|\lambda - \lambda'\|\right) \\ &\qquad = \mathbf{P}\big(\|\mathbb{G}_n^{*,\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \mathbb{G}_n^{*,\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}})\| \leq k\|\lambda - \lambda'\|\big) \geq 1 - \eta, \ \forall n \geq N. \end{aligned} \tag{E.60}$$

Moreover, by arguing as in equation (D.20), for any η there exist $0 < L < \infty$ and $N \in \mathbb{N}$ such that $\forall n \geq N$

$$\mathbf{P}\left(\sup_{\lambda' \in B^d} \left\| g_n^{\mathcal{C}}(\theta_n + \frac{\lambda'\rho}{\sqrt{n}}) \right\| \leq L\right) \geq 1 - \eta. \tag{E.61}$$

For any invertible matrix K, $\|K^{-1}\|_2 = \left(\min\{\sqrt{\alpha} : \ \alpha \ \text{ is an eigenvalue of } KK'\}\right)^{-1}$. Hence, by the proof of Lemma E.7 and the definition of $\mathcal{C}$, for any $\eta > 0$, there exist $0 < L < \infty$ and $N \in \mathbb{N}$ such that

$$\mathbf{P}\big(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2 \leq L\big) \geq 1 - \eta, \ \forall n \geq N, \tag{E.62}$$

By Horn and Johnson (1985, ch. 5.8), for any invertible matrices $K, \tilde{K}$ such that $\|\tilde{K}^{-1}(K - \tilde{K})\|_2 < 1$,

$$\|K^{-1} - \tilde{K}^{-1}\|_2 \leq \frac{\|\tilde{K}^{-1}(K - \tilde{K})\|_2}{1 - \|\tilde{K}^{-1}(K - \tilde{K})\|_2} \|\tilde{K}^{-1}\|_2. \tag{E.63}$$

By the assumption that $D_{P_n}(\theta_n) \to D$ and Assumption 4.4, for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{\lambda \in B^d} \|K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho) - K^{\mathcal{C}}\|_2 \leq \eta, \ \forall n \geq N. \tag{E.64}$$

By (E.63), the definition of the spectral norm, and the triangle inequality, for any $\eta > 0$, there exist $0 < L_1, L_2 < \infty$ and $N \in \mathbb{N}$ such that

$$\begin{aligned}
&\mathbf{P}\big(\sup_{\lambda \in B^d} \big\|\big(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)\big)^{-1}\big\|_2 \leq 2L_1\big) \\
&\quad \geq \mathbf{P}\big(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2 + \sup_{\lambda \in B^d} \|K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)^{-1} - (K^{\mathcal{C}})^{-1}\|_2 \leq 2L_1\big) \\
&\quad \geq \mathbf{P}\Bigg(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2 \leq L_1, \frac{\|\big(K^{\mathcal{C}}\big)^{-1}\|_2^2}{1 - \|\big(K^{\mathcal{C}}\big)^{-1}(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho) - K^{\mathcal{C}})\|_2} \leq L_2, \ \sup_{\lambda \in B^d} \|K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho) - K^{\mathcal{C}}\|_2 \leq \frac{L_1}{L_2}\Bigg) \\
&\quad \geq 1 - 2\eta, \ \forall n \geq N,
\end{aligned} \tag{E.65}$$

Again by applying (E.63), for any $k > 0$, there exists $N \in \mathbb{N}$ such that

$$\begin{aligned}
&\mathbf{P}\big(\big\|\big(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda))\big)^{-1} - \big(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda'))\big)^{-1}\big\|_2 \leq k\|\lambda - \lambda'\|\big) \\
&\qquad \geq \mathbf{P}\Big(\sup_{\lambda \in B^d} \big\|\big(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda))\big)^{-1}\big\|_2^2 M\rho\|\bar{\theta}(\theta_n, \lambda) - \bar{\theta}(\theta_n, \lambda')\| \leq k\|\lambda - \lambda'\|\Big) \geq 1 - \eta, \ \forall n \geq N,
\end{aligned} \tag{E.66}$$

where the first inequality follows from $\|K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda)) - K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda'))\|_2 \leq M\rho\|\bar{\theta}(\theta_n, \lambda) - \bar{\theta}(\theta_n, \lambda')\| \leq M\rho^2/\sqrt{n}\|\lambda - \lambda'\|$ by Assumption 4.4 (ii), and the last inequality follows from (E.65).

By (E.59)-(E.61) and (E.65)-(E.66), it then follows that there exists $\beta \in [0, 1)$ such that for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\mathbf{P}\left(|\phi_n(\lambda) - \phi_n(\lambda')| \leq \beta\|\lambda - \lambda'\|, \ \ \forall \lambda, \lambda' \in B^d\right) \geq 1 - \eta, \ \forall n \geq N. \tag{E.67}$$

This implies that with probability approaching 1, each $\phi_n(\cdot)$ is a contraction, and therefore by the Contraction Mapping Theorem it has a fixed point (e.g., Pata (2014, Theorem 1.3)). This in turn implies that for any $\eta > 0$ there exists a $N \in \mathbb{N}$ such that

$$\mathbf{P}\left(\exists \lambda_n^f : \lambda_n^f = \phi_n(\lambda_n^f)\right) \geq 1 - \eta, \ \forall n \geq N. \tag{E.68}$$

Next, define the mapping

$$\psi_n(\lambda) \equiv \left(K^{\mathcal{C}}\right)^{-1} g^{\mathcal{C}}. \tag{E.69}$$

This map is constant in λ and hence is uniformly continuous and a contraction with Lipschitz constant equal to zero. It therefore has $\lambda_n^{\mathcal{C}}$ as its fixed point. Moreover, by (E.58) and (E.69) arguing as in (E.59), it follows that for any $\lambda \in B^d$,

$$\begin{aligned}\|\psi_n(\lambda) - \phi_n(\lambda)\| \le \Big\|\left(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)\right)^{-1}\Big\|_2 \Big\|g^{\mathcal{C}} - g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})\Big\| \\ + \Big\|\left(K^{\mathcal{C}}\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)\right)^{-1}\Big\|_2 \|g^{\mathcal{C}}\|.\end{aligned} \tag{E.70}$$

By (E.52) and (E.56)

$$\begin{aligned}\Big\|g^{\mathcal{C}} - g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})\Big\| &\le \max_{j\in\mathcal{J}^*} | - \mathbb{Z}_j^* - c_n^*/(1+\eta_{n,j}^*) + \mathbb{G}_{n,j}^*(\theta_n + \frac{\lambda\rho}{\sqrt{n}})| \\ &\le \max_{j\in\mathcal{J}^*} |\mathbb{Z}_j^* - \mathbb{G}_{n,j}^*(\theta_n + \frac{\lambda\rho}{\sqrt{n}})| + \max_{j\in\mathcal{J}^*} |c_n^*/(1+\eta_{n,j}^*)|.\end{aligned} \tag{E.71}$$

We note that when Assumption 4.3-(II) is used, for each $j = 1, \dots, R_1$ such that $\pi_{1,j}^* = 0 = \pi_{1,j+R_1}^*$ we have that $|\tilde{\mu}_j - \mu_j| = o_{\mathcal{P}}(1)$ because $\sup_{\theta\in\Theta} |\eta_j(\theta)| = o_{\mathcal{P}}(1)$, where $\tilde{\mu}_j$ and μ_j were defined in (D.11)-(D.12) and (E.10)-(E.11) respectively. Moreover, $\mathbb{G}_{n,j}^*(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) \overset{a.s.}{\to} \mathbb{Z}^*$ and (E.49) implies $c_n^* \to 0$ so that we have

$$\sup_{\lambda\in B^d} \|g^{\mathcal{C}} - g_n^{\mathcal{C}}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})\| \overset{a.s.}{\to} 0. \tag{E.72}$$

Further, by (E.63), $D_{P_n} \to D$ and, Assumption 4.4-(ii), for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{\lambda\in B^d} \Big\|\left(K^{\mathcal{C}}\right)^{-1} - \left(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)\right)^{-1}\Big\|_2 \le \eta, \ \forall n \ge N. \tag{E.73}$$

In sum, by (E.61), (E.65), and (E.71)-(E.73), for any $\eta, \nu > 0$, there exists $N \ge \mathbb{N}$ such that

$$\mathbf{P}\left(\sup_{\lambda\in B^d} \|\psi_n(\lambda) - \phi_n(\lambda)\| < \nu\right) \ge 1 - \eta, \ \forall n \ge \mathbb{N}. \tag{E.74}$$

Hence, for a specific choice of $\nu = \kappa(1-\beta)$, where β is defined in equation (E.67), we have that $\sup_{\lambda\in B^d} \|\psi_n(\lambda) - \phi_n(\lambda)\| < \kappa(1-\beta)$ implies

$$\begin{aligned}\|\lambda_n^{\mathcal{C}} - \lambda_n^f\| &= \|\psi_n(\lambda_n^{\mathcal{C}}) - \phi_n(\lambda_n^f)\| \\ &\le \|\psi_n(\lambda_n^{\mathcal{C}}) - \phi_n(\lambda_n^{\mathcal{C}})\| + \|\phi_n(\lambda_n^{\mathcal{C}}) - \phi_n(\lambda_n^f)\| \\ &\le \kappa(1-\beta) + \beta\|\lambda_n^{\mathcal{C}} - \lambda_n^f\|\end{aligned} \tag{E.75}$$

Rearranging terms, we obtain $\|\lambda_n^{\mathcal{C}} - \lambda_n^f\| \le \kappa$. Note that by Assumptions 4.4 (i) and 4.5 (i), for any $\delta > 0$, there exists $\kappa_\delta > 0$ and $N \in \mathbb{N}$ such that

$$\mathbf{P}\Big(\sup_{\|\lambda-\lambda'\|\le\kappa_\delta} |u_{n,j,\theta_n}^*(\lambda) - u_{n,j,\theta_n}^*(\lambda')| < \delta\Big) \ge 1 - \eta, \ \forall n \ge \mathbb{N}. \tag{E.76}$$

For $\lambda_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)$, one has

$$\mathfrak{w}_j^*(\lambda_n^{\mathcal{C}}) + \delta \le 0, \ j \in \{1, \cdots, J_1\} \cap \mathcal{J}^*. \tag{E.77}$$

Hence, by (E.39), (E.49), and (E.76)-(E.77), $\|\lambda_n^{\mathcal{C}} - \lambda_n^f\| \le \kappa_{\delta/4}$, for each $j \in \{1, \cdots, J_1\} \cap \mathcal{J}^*$ we have

$$u_{n,j,\theta_n}^*(\lambda_n^f) - c_n^*(\theta_n) \le u_{n,j,\theta_n}^*(\lambda_n^{\mathcal{C}}) - c_n^*(\theta_n) + \delta/4 \le \mathfrak{w}_j^*(\lambda_n^{\mathcal{C}}) + \delta/2 \le 0. \tag{E.78}$$

For $j \in \{J_1 + 1, \cdots, 2J_2\} \cap \mathcal{J}^*$, the inequalities hold by construction given the definition of $\mathcal{C}$.

In sum, for any $\eta > 0$ there exists $\delta > 0$ and $N \in \mathbb{N}$ such that for all $n \ge N$ we have

$$\begin{aligned}\mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) = \emptyset\} \cap \{\mathfrak{W}^{*,-\delta}(0) \ne \emptyset\}\Big) &\le \mathbf{P}\Big(\nexists \lambda_n^f \in U_n^*(\theta_n, c_n^*), \exists \lambda_n^{\mathcal{C}} \in \mathfrak{W}^{*,-\delta}(0)\Big) \\ &\le \mathbf{P}\left(\left\{\sup_{\lambda \in B^d} \|\psi_n(\lambda) - \phi_n(\lambda)\| < \kappa_\delta(1-\beta) \cap A_n\right\}^c\right) \le \eta/3, \end{aligned} \tag{E.79}$$

where A^c denotes the complement of the set A, and the last inequality follows from (E.39) and (E.74). □

Lemma E.3: *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. Let $\{P_n, \theta_n\} \in \{(P,\theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ be a sequence satisfying (E.1)-(E.3). For each j, let*

$$v_{n,j,\theta_n}^I(\lambda) \equiv \mathbb{G}_{n,j}^b(\theta_n) + \rho \hat{D}_{n,j}(\theta_n)\lambda + \varphi_j^*(\hat{\xi}_{n,j}(\theta_n)), \tag{E.80}$$

$$\mathfrak{w}_j(\lambda) \equiv \mathbb{Z}_j + \rho D_j \lambda + \pi_{1,j}^*, \tag{E.81}$$

where

$$\varphi_j^*(\xi) = \begin{cases} \varphi_j(\xi) & \pi_{1,j} = 0 \\ -\infty & \pi_{1,j} < 0 \\ 0 & j = J_1 + 1, \cdots, J. \end{cases} \tag{E.82}$$

For each $c \ge 0$, define

$$V_n^I(\theta_n, c) \equiv \{\lambda \in B_{n,\rho}^d : p'\lambda = 0 \cap v_{n,j,\theta_n}^I(\lambda) \le c, j = 1, \cdots, J\}, \tag{E.83}$$

$$\mathfrak{W}(c) \equiv \big\{\lambda \in \mathfrak{B}_\rho^d : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c,\ \forall j = 1, \ldots, J\big\}. \tag{E.84}$$

We then let $c_n^I(\theta_n) \equiv \inf\{c \in \mathbb{R}_+ : P_n^*(V_n^I(\theta_n, c) \ne \emptyset) \ge 1 - \alpha\}$ and $c_{\pi^*} \equiv \inf\{c \in \mathbb{R}_+ : \Pr(\mathfrak{W}(c) \ne \emptyset) \ge 1 - \alpha\}$.

Then, (i) for any $c > 0$ and $\{\theta_n'\} \subset \Theta$ such that $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ for all n,

$$P_n^*(V_n^I(\theta_n', c) \ne \emptyset) - \Pr(\mathfrak{W}(c) \ne \emptyset) \to 0, \tag{E.85}$$

with probability approaching 1;
(ii) If $c_{\pi^*} > 0$, $c_n^I(\theta_n') \xrightarrow{P_n} c_{\pi^*}$;
(iii) For any $\{\theta_n'\} \subset \Theta$ such that $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ for all n,

$$\hat{c}_n(\theta_n') \ge c_n^I(\theta_n') + o_{P_n}(1). \tag{E.86}$$

Proof. Throughout, let $c > 0$ and let $\{\theta_n'\} \subset \Theta$ be a sequence such that $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ for all n. By Lemma E.15, in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_i\}_{i=1}^\infty$, and by Assumption 4.4 $\|\hat{D}_n(\theta_n') - D_{P_n}(\theta_n)\| \xrightarrow{p} 0$. Further, by Lemma E.5, $\hat{\xi}_{n,j}(\theta_n') \xrightarrow{P_n} \pi_{1,j}$. Therefore,

$$(\mathbb{G}_n^b(\theta_n'), \hat{D}_n(\theta_n'), \hat{\xi}_n(\theta_n')) | \{X_i\}_{i=1}^\infty \xrightarrow{d} (\mathbb{Z}, D, \pi_1). \tag{E.87}$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$. By Lemma E.17, conditional on the sample path, there exists an almost sure representation $(\tilde{\mathbb{G}}_n^b(\theta_n'), \tilde{D}_n, \tilde{\xi}_n)$ of $(\mathbb{G}_n^b(\theta_n'), \hat{D}_n(\theta_n'), \hat{\xi}_n(\theta_n'))$ defined on another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ such that $(\tilde{\mathbb{G}}_n^b(\theta_n'), \tilde{D}_n, \tilde{\xi}_n) \stackrel{d}{=} (\mathbb{G}_n^b(\theta_n'), \hat{D}_n(\theta_n'), \hat{\xi}_n(\theta_n'))$ conditional on the sample path. In particular, conditional on the sample, $(\hat{D}_n(\theta'_n), \hat{\xi}_n(\theta'_n))$ are non-stochastic. Therefore, we set $(\tilde{D}_n, \tilde{\xi}_n) = (\hat{D}_n(\theta'_n), \hat{\xi}_n(\theta'_n))$, $\tilde{\mathbf{P}} - a.s.$ The almost sure representation satisfies $(\tilde{\mathbb{G}}_n^b(\theta'_n), \tilde{D}_n, \tilde{\xi}_{n,j}) \stackrel{a.s.}{\to} (\tilde{\mathbb{Z}}, D, \pi_1)$ for almost all sample paths, where $\tilde{\mathbb{Z}} \stackrel{d}{=} \mathbb{Z}$. The almost sure representation $(\tilde{\mathbb{G}}_n^b, \tilde{D}_n, \tilde{\xi}_n)$ is defined for each sample path $x^\infty = \{x_i\}_{i=1}^\infty$, but we suppress its dependence on x^∞ for notational simplicity (see Appendix E.3 for details). Using this representation, define

$$\tilde{v}^I_{n,j,\theta'_n}(\lambda) \equiv \tilde{\mathbb{G}}^b_{n,j}(\theta'_n) + \rho \tilde{D}_n \lambda + \varphi^*_j(\tilde{\xi}_{n,j}), \tag{E.88}$$

and

$$\tilde{\mathfrak{w}}_j(\lambda) \equiv \tilde{\mathbb{Z}}_j + \rho D_j \lambda + \pi^*_{1,j}, \tag{E.89}$$

where $\tilde{\mathbb{Z}} \stackrel{d}{=} \mathbb{Z}$, and $\tilde{\mathbb{G}}^b_n(\theta'_n) \to \tilde{\mathbb{Z}}, \tilde{\mathbf{P}} - a.s.$ conditional on $\{X_i\}_{i=1}^\infty$. With this construction, one may write

$$\begin{aligned} |P^*_n(V^I_n(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset)| &= |\tilde{\mathbf{P}}(\tilde{V}^I_n(\theta'_n, c) \neq \emptyset) - \tilde{\mathbf{P}}(\tilde{\mathfrak{W}}(c) \neq \emptyset)| \\ &\leq |\tilde{\mathbf{P}}(\tilde{V}^I_n(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) + \tilde{\mathbf{P}}(\tilde{V}^I_n(\theta'_n, c) \neq \emptyset \cap \tilde{\mathfrak{W}}(c) = \emptyset)|, \end{aligned} \tag{E.90}$$

where the inequality is due to (E.33). First, we bound the first term on the right hand side of (E.90). Note that

$$\tilde{\mathbf{P}}(\tilde{V}^I_n(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) \leq \tilde{\mathbf{P}}(\tilde{V}^{I,+\delta}_n(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) + \tilde{\mathbf{P}}(\tilde{V}^{I,+\delta}_n(\theta'_n, c) \neq \emptyset \cap \tilde{V}^I_n(\theta'_n, c) = \emptyset), \tag{E.91}$$

where $\tilde{V}^{I,+\delta}_n$ is defined as

$$\tilde{V}^{I,+\delta}_n \equiv \Big\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap \tilde{v}^I_{n,j,\theta'_n}(\lambda) \leq c + \delta, j \in \mathcal{J}^*\Big\}. \tag{E.92}$$

Let

$$A_n \equiv \Big\{\tilde{\omega} \in \tilde{\Omega} : \sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} |\tilde{v}^I_{n,j,\theta'_n}(\lambda) - \tilde{\mathfrak{w}}_j(\lambda)| \geq \delta\Big\}. \tag{E.93}$$

Let

$$E \equiv \{\{x_i\}_{i=1}^\infty : \|\hat{D}_n(\theta'_n) - D\| < \eta, \max_{j \in \mathcal{J}^*} |\varphi^*_j(\hat{\xi}_{n,j}(\theta'_n)) - \pi^*_{1,j}| < \eta\}. \tag{E.94}$$

Note that, $P_n(E) \geq 1 - \eta$ for all n sufficiently large by Assumption 4.4 and Lemma E.5. On E, we therefore have $\|\tilde{D}_n - D\| < \eta$ and $\max_{j \in \mathcal{J}^*} |\tilde{\xi}_{n,j} - \pi^*_{1,j}| < \eta$, $\tilde{\mathbf{P}} - a.s.$ Below, we condition on $\{X_i\}_{i=1}^\infty \in E$. For any $j \in \mathcal{J}^*$,

$$|\tilde{v}^I_{n,j,\theta'_n}(\lambda) - \tilde{\mathfrak{w}}_j(\lambda)| \leq |\tilde{\mathbb{G}}^b_{n,j}(\theta'_n) - \tilde{\mathbb{Z}}_j| + \rho\|\tilde{D}_{j,n} - D_j\|\|\lambda\| + |\varphi^*_j(\tilde{\xi}_{n,j}) - \pi^*_{1,j}| \leq (2 + \rho)\eta, \tag{E.95}$$

uniformly in $\lambda \in B^d$, where we used $\tilde{\mathbb{G}}^b_n \to \tilde{\mathbb{Z}}, \tilde{\mathbf{P}} - a.s.$ Since η can be chosen arbitrarily small, this in turn implies

$$\tilde{\mathbf{P}}\big(A_n\big) < \eta/2,$$

for all n sufficiently large. Note also that $\sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} |\tilde{v}^I_{n,j,\theta'_n}(\lambda) - \tilde{\mathfrak{w}}_j(\lambda)| < \delta$ implies $\tilde{\mathfrak{W}}(c) \subseteq \tilde{V}^{I,+\delta}_n(\theta'_n, c)$, and hence A^c_n is a subset of

$$L_n \equiv \Big\{\tilde{\omega} \in \tilde{\Omega} : \tilde{\mathfrak{W}}(c) \subseteq \tilde{V}^{I,+\delta}_n(\theta'_n, c)\Big\}. \tag{E.96}$$

Using this,

$$\tilde{\mathbf{P}}(\tilde{V}^{I,+\delta}_n(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) \leq \tilde{\mathbf{P}}(\tilde{\mathfrak{W}}(c) \not\subseteq \tilde{V}^{I,+\delta}_n(\theta'_n, c)) = \tilde{\mathbf{P}}(L^c_n) \leq \tilde{\mathbf{P}}(A_n) < \eta/2, \tag{E.97}$$

for all n sufficiently large. Also, by Lemma E.6,

$$\tilde{\mathbf{P}}(\tilde{V}_n^{I,+\delta}(\theta_n',c)\neq\emptyset\cap\tilde{V}_n^{I}(\theta_n',c)=\emptyset)<\eta/2, \tag{E.98}$$

for all n sufficiently large.

Combining (E.91), (E.93), (E.97), (E.98), and using $P_n(E)\geq 1-\eta$ for all n, we have

$$\int_E \tilde{\mathbf{P}}(\tilde{V}_n^I(\theta_n',c)=\emptyset\cap\tilde{\mathfrak{W}}(c)\neq\emptyset)dP_n+\int_{E^c}\tilde{\mathbf{P}}(\tilde{V}_n^I(\theta_n',c)=\emptyset\cap\tilde{\mathfrak{W}}(c)\neq\emptyset)dP_n\leq\eta(1-\eta)+\eta\leq 2\eta. \tag{E.99}$$

The second term of the right hand side of (E.90) can be bounded similarly. Therefore, $|P^*(V_n^I(\theta_n',c)\neq\emptyset)-\Pr(\mathfrak{W}(c)\neq\emptyset)|\to 0$ with probability (under P_n) approaching 1. This establishes the first claim.

(ii) By Part (i), for $c>0$, we have

$$P_n^*(V_n^I(\theta_n',c)\neq\emptyset)-\Pr(\mathfrak{W}(c)\neq\emptyset)\to 0. \tag{E.100}$$

Fix $c>0$, and set

$$g_j=\begin{cases} c-\mathbb{Z}_j, & j=1,\dots,J,\\ 1, & j=J+1,\dots,J+2d,\\ 0, & j=J+2d+1,J+2d+2.\end{cases} \tag{E.101}$$

Mimic the argument following (E.137). Then, this yields

$$|\Pr(\mathfrak{W}(c)\neq\emptyset)-\Pr(\mathfrak{W}(c-\delta)\neq\emptyset)|=\Pr(\{\mathfrak{W}(c)\neq\emptyset\}\cap\{\mathfrak{W}(c-\delta)=\emptyset\})\leq\eta, \tag{E.102}$$

$$|\Pr(\mathfrak{W}(c+\delta)\neq\emptyset)-\Pr(\mathfrak{W}(c)\neq\emptyset)|=\Pr(\{\mathfrak{W}(c+\delta)\neq\emptyset\}\cap\{\mathfrak{W}(c)=\emptyset\})\leq\eta, \tag{E.103}$$

which therefore ensures that $c\mapsto\Pr(\mathfrak{W}(c)\neq\emptyset)$ is continuous at $c>0$.

Next, we show $c\mapsto\Pr(\mathfrak{W}(c)\neq\emptyset)$ is strictly increasing at any $c>0$. For this, consider $c>0$ and $c-\delta>0$ for $\delta>0$. Define the J vector e to have elements $e_j=c-\mathbb{Z}_j,\ j=1,\dots,J$. Suppose for simplicity that $\mathcal{J}^*$ contains the first J^* inequality constraints. Let $e^{[1:J^*]}$ denote the subvector of e that only contains elements corresponding to $j\in\mathcal{J}^*$, define $D^{[1:J^*,:]}$ correspondingly, and write

$$K=\begin{bmatrix} D^{[1:J^*,:]}\\ I_d\\ -I_d\\ p'\\ -p'\end{bmatrix},\quad g=\begin{bmatrix} e^{[1:J^*]}\\ \rho\cdot\mathbf{1}_d\\ \rho\cdot\mathbf{1}_d\\ 0\\ 0\end{bmatrix},\quad \tau=\begin{bmatrix}\mathbf{1}_{J^*}\\ \mathbf{0}_d\\ \mathbf{0}_d\\ 0\\ 0\end{bmatrix}. \tag{E.104}$$

By Farkas' lemma (Rockafellar, 1970, Theorem 22.1) and arguing as in (E.142),

$$\Pr(\{\mathfrak{W}(c)\neq\emptyset\}\cap\{\mathfrak{W}(c-\delta)=\emptyset\})=\Pr(\{\mu'g\geq 0,\forall\mu\in\mathcal{M}\}\cap\{\mu'(g-\delta\tau)<0,\exists\mu\in\mathcal{M}\}), \tag{E.105}$$

where $\mathcal{M}=\{\mu\in\mathbb{R}_+^{J^*+2d+2}:\mu'K=0\}$. By Minkowski-Weyl's theorem (Rockafellar and Wets, 2005, Theorem 3.52), there exists $\{\nu^t\in\mathcal{M},t=1,\cdots,T\}$, for which one may write

$$\mathcal{M}=\{\mu:\mu=b\sum_{t=1}^T a_t\nu^t,b>0,a_t\geq 0,\sum_{t=1}^T a_t=1\}. \tag{E.106}$$

This implies

$$\mu' g \geq 0,\ \forall \mu \in \mathcal{M} \ \Leftrightarrow\ \nu^{t\prime} g \geq 0,\ \forall t \in \{1, \cdots, T\} \tag{E.107}$$

$$\mu'(g - \delta\tau) < 0,\ \exists \mu \in \mathcal{M} \ \Leftrightarrow\ \nu^{t\prime} g < \delta \nu^{t\prime}\tau,\ \exists t \in \{1, \cdots, T\}. \tag{E.108}$$

Hence,

$$\Pr\left(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\}\right) = \Pr\left(0 \leq \nu^{s\prime} g,\ 0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau,\ \forall s, \exists t\right) \tag{E.109}$$

Note that by (E.104), for each $s \in \{1, \cdots, T\}$,

$$\nu^{s\prime} g = \nu^{s,[1:J^*]\prime}(c1_{\mathcal{J}^*} - \mathbb{Z}_{\mathcal{J}^*}) + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,[j]}, \tag{E.110}$$

$$\nu^{s\prime}\tau = \sum_{j=1}^{J^*} \nu^{s,[j]}. \tag{E.111}$$

For each $s \in \{1, \cdots, T\}$, let

$$h_s^U \equiv c \sum_{j=1}^{J^*} \nu^{s,[j]} + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,[j]} \tag{E.112}$$

$$h_s^L \equiv (c - \delta) \sum_{j=1}^{J^*} \nu^{s,[j]}, \tag{E.113}$$

where $0 \leq h_s^L < h_s^U$ for all $s \in \{1, \cdots, T\}$ due to $0 < c - \delta < c$ and $\nu^s \in \mathbb{R}_+^{J^*+2d+2}$. One may therefore rewrite the probability on the right hand side of (E.109) as

$$\Pr\left(0 \leq \nu^{s\prime} g,\ 0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau,\ \forall s, \exists t\right) = \Pr\left(\nu^{s,[1:J^*]\prime}\mathbb{Z}_{\mathcal{J}^*} \leq h_s^U, h_t^L < \nu^{t,[1:J^*]\prime}\mathbb{Z}_{\mathcal{J}^*} \leq h_t^U\ \forall s, \exists t\right) > 0, \tag{E.114}$$

where the last inequality follows because $\mathbb{Z}_{\mathcal{J}^*}$'s correlation matrix Ω has an eigenvalue bounded away from 0 by Assumption 4.3. By (E.105), (E.109), and (E.114), $c \mapsto \Pr(\mathfrak{W}(c) \neq \emptyset)$ is strictly increasing at any $c > 0$.

Suppose that $c_{\pi^*} > 0$, then arguing as in Lemma 5.(i) of Andrews and Guggenberger (2010), we obtain $c_n^I(\theta_n') \stackrel{P_n}{\to} c_{\pi^*}$.

(iii) Begin with observing that one can equivalently express $\hat{c}_n$ (originally defined in (3.5)) as $\hat{c}_n(\theta) = \inf\{c \in \mathbb{R}_+ : P_n^*(V_n^b(\theta, c) \neq \emptyset) \geq 1 - \alpha\}$.

Suppose first that Assumption 4.3-(I) holds. In this case, there are no paired inequalities, and V_n^I differs from V_n^b only in terms of the function φ_j^* in (E.82) used in place of the GMS function φ_j. In particular, $\varphi_j^*(\xi) \leq \varphi_j(\xi)$ for any j and ξ, and therefore $\hat{c}_n(\theta_n) \geq c_n^I(\theta_n)$ by construction.

Next, suppose Assumption 4.3-(II) holds and $V_n^I(\theta_n', c)$ is defined with hard threshold GMS as in equation (3.3), i.e. with GMS function φ^1 in AS. The only case that might create concern is one in which

$$\pi_{1,j} \in [-1, 0) \text{ and } \pi_{1,j+R_1} = 0. \tag{E.115}$$

In this case, only the $j + R_1$-th inequality binds in the limit, but with probability approaching 1, GMS selects both of the pair. Therefore, we have

$$\pi^*_{1,j} = -\infty, \text{ and } \pi^*_{1,j+R_1} = 0, \tag{E.116}$$

$$\varphi_j(\hat{\xi}_{n,j}(\theta'_n)) = 0, \text{ and } \varphi_{j+R_1}(\hat{\xi}_{n,j+R_1}(\theta'_n)) = 0, \tag{E.117}$$

so that in $V_n^I(\theta'_n, c)$, inequality $j + R_1$, which is

$$\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda \le c, \tag{E.118}$$

is replaced with inequality

$$-\mathbb{G}^b_{n,j}(\theta'_n) - \rho\hat{D}_{n,j}(\theta'_n)\lambda \le c, \tag{E.119}$$

as explained in Section 4.1. In this case, $\hat{c}_n(\theta_n) \ge c_n^I(\theta_n)$ is not guaranteed in finite sample. However, let v_n^{IP} be as in (E.80) but replacing $j + R_1$-th component $\mathbb{G}^b_{n,j+R_1}(\theta_n) + \hat{D}_{n,j+R_1}(\theta_n)\lambda + \varphi^*_{j+R_1}(\hat{\xi}_{n,j+R_1}(\theta_n))$ with $-\mathbb{G}^b_{n,j}(\theta_n) - \hat{D}_{n,j}(\theta_n)\lambda - \varphi^*_j(\hat{\xi}_{n,j}(\theta_n))$. Define V_n^{IP} as in (E.83) but replacing v_n^I with v_n^{IP}. Define $c_n^{IP}(\theta_n) \equiv \inf\{c \in \mathbb{R}_+ : P^*(V_n^{IP}(\theta_n, c)) \ge 1-\alpha\}$. By construction, $\hat{c}_n(\theta'_n) \ge c_n^{IP}(\theta'_n)$ for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$. Therefore, it suffices to show that $c_n^{IP}(\theta'_n) - c_n^I(\theta'_n) \overset{P_n}{\to} 0$. For this, note that Lemma E.9-(3) establishes

$$\sup_{\lambda \in B^d_{n,\rho}} \|\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda + \mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\| = o_{P^*}(1), \tag{E.120}$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$. Therefore, replacing the $j + R_1$-th inequality with the j-th inequality in V_n^{IP} is asymptotically negligible. Mimicking the arguments in Parts (i) and (ii) then yields

$$c_n^{IP}(\theta'_n) \overset{P_n}{\to} c_{\pi^*}. \tag{E.121}$$

This therefore ensures $c_n^{IP}(\theta'_n) - c_n^I(\theta'_n) \overset{P_n}{\to} 0$.

If the set $V_n^I(\theta'_n, c)$ is defined with a GMS function satisfying Assumption 4.2 and continuous in its argument, we can mimic the above argument using the replacements in (E.12)-(E.13) with $\hat{\mu}_{n,j+R_1}$ as defined in (E.14) and $\hat{\mu}_{n,j}(\theta'_n)$ as in (E.15). Then when both $\pi_j \in (-\infty, 0]$ and $\pi_{j+R_1} \in (-\infty, 0]$ we have:

$$\begin{aligned}\Delta(\mu,\hat{\mu}) \equiv \Big\|&\hat{\mu}_{n,j}(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\} - \hat{\mu}_{n,j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\}\\ &-\mu_j(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\} + \mu_{j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\}\Big\| = o_{\mathcal{P}}(1),\end{aligned}$$

where μ_j, μ_{j+R_1} are defined in equations (E.10)-(E.11) for $\theta \in \theta_n + (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$. Replacing $\hat{\mu}_{n,j} = 1 - \hat{\mu}_{n,j+R_1}$ and $\mu_j = 1 - \mu_{j+R_1}$ in the definition of $\Delta(\mu, \hat{\mu})$, we have

$$\Delta(\mu,\hat{\mu}) \le \big|\hat{\mu}_{n,j+R_1}(\theta'_n) - \mu_{j+R_1}(\theta'_n)\big|\big\|\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\} + \{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\}\big\|. \tag{E.122}$$

If both $\pi_j \in (-\infty, 0], \pi_{j+R_1} \in (-\infty, 0]$, the result follows by the fact that $\lambda \in B^d_{n,\rho}$ and $\hat{\mu}_{n,j}, \hat{\mu}_{n,j+R_1}, \mu_j, \mu_{j+R_1}$ are bounded in $[0, 1]$, by Lemma E.9-(3)-(4), and by Assumption 4.4-(i). The rest of the argument follows similarly as for the case of hard-threshold GMS. □

LEMMA E.4: *Let Assumptions 4.1, 4.2, 4.4, and 4.5 hold. Let* (P_n, θ_n) *be the sequence satisfying* (E.1)-(E.3)*, let* $\mathcal{J}^*$ *be defined as in* (E.29)*, and assume that* $\mathcal{J}^* \neq \emptyset$*. Then, for any* $\varepsilon, \eta > 0$ *and* $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$*, there* *exists* $N' \in \mathbb{N}$ *and* $N'' \in \mathbb{N}$ *such that for all* $n \geq \max\{N', N''\}$,

$$\mathbf{P}\left(\sup_{\lambda \in B^d} \left| \max_{j=1,\cdots,J} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j=1,\cdots,J} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*}) \right| \geq \varepsilon \right) < \eta, \tag{E.123}$$

$$\tilde{\mathbf{P}}\left(\sup_{\lambda \in B^d} \left| \max_{j=1,\cdots,J} \tilde{\mathfrak{w}}_j(\lambda) - \max_{j=1,\cdots,J} \tilde{v}^I_{n,j,\theta'_n(\lambda)} \right| \geq \varepsilon \right) < \eta, \text{ w.p.1}, \tag{E.124}$$

where the functions $u^*_n, \mathfrak{w}^*, \tilde{v}_n, \tilde{\mathfrak{w}}$ *are defined in equations* (E.24),(E.25), (E.88), *and* (E.89).

Proof. We first establish (E.123). By definition, $\pi^*_{1,j} = -\infty$ for all $j \notin \mathcal{J}^*$ and therefore

$$\mathbf{P}\Big(\sup_{\lambda \in B^d} | \max_{j=1,\cdots,J} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j=1,\cdots,J} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \geq \varepsilon\Big)$$
$$= \mathbf{P}\Big(\sup_{\lambda \in B^d} | \max_{j \in \mathcal{J}^*} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j \in \mathcal{J}^*} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \geq \varepsilon\Big). \tag{E.125}$$

Hence, for the conclusion of the lemma, it suffices to show, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big(\sup_{\lambda \in B^d} | \max_{j \in \mathcal{J}^*} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j \in \mathcal{J}^*} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \geq \varepsilon\Big) = 0.$$

For each $\lambda \in \mathbb{R}^d$, define $r_{n,j,\theta_n}(\lambda) \equiv (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - (\mathfrak{w}^*_j(\lambda) - c_n)$. Using the fact that $\pi^*_{1,j} = 0$ for $j \in \mathcal{J}^*$, and the triangle and Cauchy-Schwarz inequalities, for any $\lambda \in B^d \cap \frac{\sqrt{n}}{\rho}(\Theta - \theta_n)$ and $j \in \mathcal{J}^*$, we have

$$\begin{aligned}
|r_{n,j,\theta_n}(\lambda)| &\leq |\mathbb{G}^*_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j| + \rho\|D_{P_n,j}(\bar{\theta}_n) - D_j\|\|\lambda\| \\
&\quad + |\mathbb{G}^*_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda|\eta^*_{n,j} + |c^*_n - c_{\pi^*}| \\
&= |\mathbb{G}^*_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j| + o(1) + \{O_{\mathcal{P}}(1) + O(1)\})\eta^*_{n,j} + o_{\mathcal{P}}(1) \\
&= o_{\mathcal{P}}(1)
\end{aligned} \tag{E.126}$$

where the first equality follows from $\|\lambda\| \leq \sqrt{d}$, $D_{P_n}(\bar{\theta}_n) \to D$ due to $D_{P_n}(\theta_n) \to D$, Assumption 4.4-(ii), and $\bar{\theta}_n$ being a mean value between θ_n and $\theta_n + \lambda\rho/\sqrt{n}$. We also note that $\|\mathbb{G}_{n,j}(\theta + \lambda/\sqrt{n})\| = O_{\mathcal{P}}(1)$, $\|D_{P,j}(\theta)\|$ being uniformly bounded for $\theta \in \Theta_I(P)$ (Assumption 4.4-(i)), and $c^*_n \overset{a.s.}{\to} c_{\pi^*}$. The last equality follows from $\mathbb{G}^*_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j \overset{a.s.}{\to} 0$ and $\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| = o_{\mathcal{P}}(1)$ by Lemma E.10.

We note that when paired inequalities are merged, for each $j = 1, \ldots, R_1$ such that $\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1}$ we have that $|\tilde{\mu}_j - \mu_j| = o_{\mathcal{P}}(1)$ because $\sup_{\theta\in\Theta} |\eta_j(\theta)| = o_{\mathcal{P}}(1)$, where $\tilde{\mu}_j$ and μ_j were defined in (D.11)-(D.12) and (E.10)-(E.11) respectively.

By (E.126) and the fact that $j \in \mathcal{J}^*$, we have

$$\sup_{\lambda \in B^d} | \max_{j \in \mathcal{J}^*} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j \in \mathcal{J}^*} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \leq \sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} |r_{n,j,\theta_n}(\lambda)| = o_{\mathcal{P}}(1). \tag{E.127}$$

The conclusion of the lemma then follows from (E.125) and (E.127).

The result in (E.124) follows from similar arguments. ∎

Lemma E.5: *Let Assumptions 4.1, 4.2, 4.4, and 4.5 hold. Given a sequence* $\{Q_n, \vartheta_n\} \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ *such that* $\lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\vartheta_n)$ *exists for each* $j = 1, \ldots, J$, *let* $\chi_j(\{Q_n, \vartheta_n\})$ *be a function of the* *sequence* $\{Q_n, \vartheta_n\}$ *defined as*

$$\chi_j(\{Q_n, \vartheta_n\}) \equiv \begin{cases} 0, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\vartheta_n) = 0, \\ -\infty, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\vartheta_n) < 0. \end{cases} \tag{E.128}$$

Then for any $\theta'_n \in \theta_n + \frac{\rho}{\sqrt{n}}B^d$ *for all* n*, one has: (i)* $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) - \kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta'_n) = o(1)$*; (ii)* $\chi(\{P_n, \theta_n\}) = \chi(\{P_n, \theta'_n\}) = \pi^*_{1,j}$*; and (iii)* $\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}(\theta'_n)}{\hat{\sigma}_{n,j}(\theta'_n)} - \kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\sigma_{P_n,j}(\theta'_n)} = o_{\mathcal{P}}(1)$.

Proof. For (i), the mean value theorem yields

$$\sup_{P\in\mathcal{P}} \sup_{\theta\in\Theta_I(P),\theta'\in\theta+\rho/\sqrt{n}B^d} \left| \frac{\sqrt{n}E_P(m_j(X,\theta))}{\kappa_n\sigma_{P,j}(\theta)} - \frac{\sqrt{n}E_P(m_j(X,\theta'))}{\kappa_n\sigma_{P,j}(\theta')} \right| \\ \leq \sup_{P\in\mathcal{P}} \sup_{\theta\in\Theta_I(P),\theta'\in\theta+\rho/\sqrt{n}B^d} \frac{\sqrt{n}\|D_{P,j}(\tilde{\theta})\|\|\theta'-\theta\|}{\kappa_n} = o(1), \tag{E.129}$$

where $\tilde{\theta}$ represents a mean value that lies componentwise between θ and θ' and where we used the fact that $D_{P,j}(\theta)$ is Lipschitz continuous and $\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P)}\|D_{P,j}(\theta)\| \leq \bar{M}$. Result (ii) then follows immediately from (E.128). For (iii), note that

$$\sup_{\theta'_n\in\theta_n+\rho/\sqrt{n}B^d} \left|\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}(\theta'_n)}{\hat{\sigma}_{n,j}(\theta'_n)} - \kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\sigma_{P_n,j}(\theta'_n)}\right| \\ \leq \sup_{\theta'_n\in\theta_n+\rho/\sqrt{n}B^d} \left|\kappa_n^{-1}\frac{\sqrt{n}(\bar{m}_{n,j}(\theta'_n) - E_{P_n}[m_j(X_i,\theta'_n)])}{\sigma_{n,j}(\theta'_n)}(1+\eta_{n,j}(\theta'_n)) + \kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\sigma_{P_n,j}(\theta'_n)}\eta_{n,j}(\theta'_n)\right| \\ \leq \sup_{\theta'_n\in\theta_n+\rho/\sqrt{n}B^d} |\kappa_n^{-1}\mathbb{G}_n(\theta'_n)(1+\eta_{n,j}(\theta'_n))| + \left|\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\kappa_n\sigma_{P_n,j}(\theta'_n)}\eta_{n,j}(\theta'_n)\right| = o_{\mathcal{P}}(1), \tag{E.130}$$

where the last equality follows from $\sup_{\theta\in\Theta}|\mathbb{G}_n(\theta)| = O_{\mathcal{P}}(1)$ due to asymptotic tightness of $\{\mathbb{G}_n\}$ (uniformly in P) by Lemma D.1 in Bugni, Canay, and Shi (2015), Theorem 3.6.1 and Lemma 1.3.8 in van der Vaart and Wellner (2000), and $\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| = o_{\mathcal{P}}(1)$ by Lemma E.10-(i). □

LEMMA E.6: *Let Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. For any* $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$,

(i) For any $\eta > 0$*, there exist* $\delta > 0$ *such that*

$$\sup_{c\geq 0} \Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}^{-\delta}(c) = \emptyset\}) < \eta. \tag{E.131}$$

Moreover, for any $\eta > 0$*, there exist* $\delta > 0$ *and* $N \in \mathbb{N}$ *such that*

$$\sup_{c\geq 0} P^*_n(\{V^I_n(\theta'_n, c) \neq \emptyset\} \cap \{V^{I,-\delta}_n(\theta'_n, c) = \emptyset\}) < \eta, \ \forall n \geq N. \tag{E.132}$$

(ii) Fix $\underline{c} > 0$ *and redefine*

$$\mathfrak{W}^{-\delta}(c) \equiv \{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \leq c - \delta, \ \forall j = 1, \dots, J\}, \tag{E.133}$$

and

$$V^{I,-\delta}_n(\theta'_n, c) \equiv \{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap v^I_{n,j,\theta'_n}(\lambda) \leq c - \delta, \ \forall j = 1, \dots, J\}. \tag{E.134}$$

Then for any $\eta > 0$, there exists $\delta > 0$ such that

$$\sup_{c \geq \underline{c}} \Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}^{-\delta}(c) = \emptyset\}) < \eta. \tag{E.135}$$

with $\mathfrak{W}^{-\delta}(c)$ defined in (E.133). Moreover, for any $\eta > 0$, there exist $\delta > 0$ and $N \in \mathbb{N}$ such that

$$\sup_{c \geq \underline{c}} P_n^*(\{V_n^I(\theta_n', c) \neq \emptyset\} \cap \{V_n^{I,-\delta}(\theta_n', c) = \emptyset\}) < \eta, \ \forall n \geq N, \tag{E.136}$$

with $V_n^{-\delta}(\theta_n', c)$ defined in (E.134).

Proof. We first show (E.131). If $\mathcal{J}^* = \emptyset$, with $\mathcal{J}^*$ as defined in (E.29), then the result is immediate. Assume then that $\mathcal{J}^* \neq \emptyset$. Any inequality indexed by $j \notin \mathcal{J}^*$ is satisfied with probability approaching one by similar arguments as in (D.20) (both with c and with $c - \delta$). Hence, one could argue for sets $\mathfrak{W}(c), \mathfrak{W}^{-\delta}(c)$ defined as in equations (E.16) and (E.17) but with $j \in \mathcal{J}^*$. To keep the notation simple, below we argue as if all $j = 1, \ldots, J$ belong to $\mathcal{J}^*$. Let $c \geq 0$ be given. Let g be a $J + 2d + 2$ vector with entries

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j = 1, \ldots, J, \\ 1, & j = J+1, \ldots, J+2d, \\ 0, & j = J+2d+1, J+2d+2, \end{cases} \tag{E.137}$$

recalling that $\pi_{1,j}^* = 0$ for $j = J_1 + 1, \cdots, J$. Let τ be a $(J + 2d + 2)$ vector with entries

$$\tau_j = \begin{cases} 1, & j = 1, \ldots, J_1, \\ 0, & j = J_1 + 1, \ldots, J + 2d + 2. \end{cases} \tag{E.138}$$

Then we can express the sets of interest as

$$\mathfrak{W}(c) = \{\lambda : K\lambda \leq g\}, \tag{E.139}$$

$$\mathfrak{W}^{-\delta}(c) = \{\lambda : K\lambda \leq g - \delta\tau\}. \tag{E.140}$$

By Farkas' Lemma, e.g. Rockafellar (1970, Theorem 22.1), a solution to the system of linear inequalities in (E.139) exists if and only if for all $\mu \in \mathbb{R}_+^{J+2d+2}$ such that $\mu' K = 0$, one has $\mu' g \geq 0$. Similarly, a solution to the system of linear inequalities in (E.140) exists if and only if for all $\mu \in \mathbb{R}^{J+2d+2}$ such that $\mu' K = 0$, one has $\mu'(g - \delta\tau) \geq 0$. Define

$$\mathcal{M} \equiv \{\mu \in \mathbb{R}_+^{J+2d+2} : \mu' K = 0\}. \tag{E.141}$$

Then, one may write

$$\begin{aligned} &\Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{W^{-\delta}(\theta_n', c) = \emptyset\}) \\ &\quad = \Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\}) \\ &\quad = \Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu' g < \delta\mu'\tau, \exists \mu \in \mathcal{M}\}). \end{aligned} \tag{E.142}$$

Note that the set $\mathcal{M}$ is a non-stochastic polyhedral cone which may change with n. By Minkowski-Weyl's theorem (see, e.g. Rockafellar and Wets (2005, Theorem 3.52)), for each n there exist $\{\nu^t \in \mathcal{M}, t = 1, \cdots, T\}$, with $T < \infty$ a constant that depends only on J and d, such that any $\mu \in \mathcal{M}$ can be represented as

$$\mu = b \sum_{t=1}^{T} a_t \nu^t, \tag{E.143}$$

where $b > 0$ and $a_t \geq 0,\ t = 1, \ldots, T,\ \sum_{t=1}^{T} a_t = 1$. Hence, if $\mu \in \mathcal{M}$ satisfies $\mu' g < \delta \mu' \tau$, denoting $\nu^{t\prime}$ the transpose of vector ν^t, we have

$$\sum_{t=1}^{T} a_t \nu^{t\prime} g < \delta \sum_{t=1}^{T} a_t \nu^{t\prime} \tau. \tag{E.144}$$

However, due to $a_t \geq 0, \forall t$ and $\nu^t \in \mathcal{M}$, this means $\nu^{t\prime} g < \delta \nu^{t\prime} \tau$ for some $t \in \{1, \ldots, T\}$. Furthermore, since $\nu^t \in \mathcal{M}$, we have $0 \leq \nu^{t\prime} g$. Therefore,

$$\begin{aligned} &\Pr\left(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu' g < \delta \mu' \tau, \exists \mu \in \mathcal{M}\}\right) \\ &\qquad \leq \Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \exists t \in \{1, \cdots, T\}\right) \leq \sum_{t=1}^{T} \Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right). \end{aligned} \tag{E.145}$$

Case 1. Consider first any $t = 1, \ldots, T$ such that ν^t assigns positive weight only to constraints in $\{J+1, \ldots, J+2d+2\}$. Then

$$\nu^{t\prime} g = \sum_{j=J+1}^{J+2d} \nu_j^t,$$
$$\delta \nu^{t\prime} \tau = \delta \sum_{j=J+1}^{J+2d+2} \nu_j^t \tau_j = 0,$$

where the last equality follows by (E.138). Therefore $\Pr(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau) = 0$.

Case 2. Consider now any $t = 1, \ldots, T$ such that ν^t assigns positive weight also to constraints in $\{1, \ldots, J\}$. Recall that indices $j = J_1 + 1, \ldots, J_1 + 2J_2$ correspond to moment equalities, each of which is written as two moment inequalities, therefore yielding a total of $2J_2$ inequalities with $D_{j+J_2} = -D_j$ for $j = J_1 + 1, \ldots, J_1 + J_2$, and:

$$g = \begin{cases} c - \mathbb{Z}_j & j = J_1 + 1, \ldots, J_1 + J_2, \\ c + \mathbb{Z}_{j-J_2} & j = J_1 + J_2 + 1, \ldots, J. \end{cases} \tag{E.146}$$

For each ν^t, (E.146) implies

$$\sum_{j=J_1+1}^{J_1+2J_2} \nu_j^t g_j = c \sum_{j=J_1+1}^{J_1+2J_2} \nu_j^t + \sum_{j=J_1+1}^{J_1+J_2} (\nu_j^t - \nu_{j+J_2}^t) \mathbb{Z}_j. \tag{E.147}$$

For each $j = 1, \cdots, J_1 + J_2$, define

$$\tilde{\nu}_j^t \equiv \begin{cases} \nu_j^t & j = 1, \cdots, J_1 \\ \nu_j^t - \nu_{j+J_2}^t & j = J_1 + 1, \cdots, J_1 + J_2. \end{cases}. \tag{E.148}$$

We then let $\tilde{\nu}^t \equiv (\tilde{\nu}_{n,1}^t, \cdots, \tilde{\nu}_{n,J_1+J_2}^t)'$ and have

$$\nu^{t\prime} g = \sum_{j=1}^{J_1+J_2} \tilde{\nu}_j^t \mathbb{Z}_j + c \sum_{j=1}^{J} \nu_j^t + \sum_{j=J+1}^{J+2d} \nu_j^t. \tag{E.149}$$

Case 2-a. Suppose $\tilde{\nu}^t \neq 0$. Then, by (E.149), $\frac{\nu^{t\prime} g}{\nu^{t\prime}\tau}$ is a normal random variable with variance $(\tilde{\nu}^{t\prime}\tau)^{-2}\tilde{\nu}'^t\Omega\tilde{\nu}^t$. By Assumption 4.3, there exists a constant $\omega > 0$ such that the smallest eigenvalue of Ω is bounded from below by ω for all θ'_n. Hence, letting $\|\cdot\|_p$ denote the p-norm in $\mathbb{R}^{J+2d+2}$, we have

$$\frac{\tilde{\nu}'^t\Omega\tilde{\nu}^t}{(\tilde{\nu}^{t\prime}\tau)^2} \geq \frac{\omega\|\tilde{\nu}^t\|_2^2}{(J+2d+2)^2\|\tilde{\nu}^t\|_2^2} \geq \frac{\omega}{(J+2d+2)^2}. \tag{E.150}$$

Therefore, the variance of the normal random variable in (E.145) is uniformly bounded away from 0, which in turn allows one to find $\delta > 0$ such that $\Pr(0 \leq \frac{\nu^{t\prime} g}{\nu^{t\prime}\tau} < \delta) \leq \eta/T$.

Case 2-b. Next, consider the case $\tilde{\nu}^t = 0$. Because we are in the case that ν^t assigns positive weight also to constraints in $\{1, \ldots, J\}$, this must be because $\nu_j^t = 0$ for all $j = 1, \cdots, J_1$ and $\nu_j^t = \nu_{j+J_2}^t$ for all $j = J_1 + 1, \cdots, J_1 + J_2$, while $\nu_j^t \neq 0$ for some $j = J_1 + 1, \cdots, J_1 + J_2$. Then we have $\sum_{j=1}^J \nu_j^t g \geq 0$, and $\sum_{j=1}^J \nu_j^t \tau_j = 0$ because $\tau_j = 0$ for each $j = J_1 + 1, \ldots, J$. Hence, the argument for the case that ν^t assigns positive weight only to constraints in $\{J+1, \ldots, J+2d+2\}$ applies and again $\Pr(0 \leq \nu^{t\prime} g < \delta\nu^{t\prime}\tau) = 0$. This establishes equation (E.131).

To see why equation (E.132) holds, observe that the bootstrap distribution is conditional on $X_1, \ldots, X_n$. Therefore, the matrix $\hat{K}_n$, defined as the matrix in equation (E.57) but with $\hat{D}_n$ replacing D_P, can be treated as nonstochastic. This implies that the set $\hat{\mathcal{M}}_n$, defined as the set in equation (E.141) but with $\hat{K}_n$ replacing K, can be treated as nonstochastic as well.

By an application of Lemma D.2.8 in Bugni, Canay, and Shi (2015) together with Lemma E.17 (through an argument similar to that following equation (E.87)), $\mathbb{G}_n^b \xrightarrow{d} \mathbb{G}_P$ in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_1, \cdots, X_n\}$, and by Assumption 4.4 $\hat{D}_n(\theta'_n) \xrightarrow{P_n} D$, for almost all sample paths. Set

$$g_{P_n,j}(\theta'_n) = \begin{cases} c - \varphi_j^*(\xi_{n,j}(\theta'_n)) - \mathbb{G}_{n,j}^b(\theta'_n), & j = 1, \ldots, J, \\ 1, & j = J+1, \ldots, J+2d, \\ 0, & j = J+2d+1, J+2d+2, \end{cases} \tag{E.151}$$

and note that $|\varphi_j^*(\xi_{n,j}(\theta'_n))| < \eta$ for all $j \in \mathcal{J}^*$, and $\mathbb{G}_{n,j}^b(\theta'_n)|\{X_i\}_{i=1}^\infty \xrightarrow{d} N(0, \Omega)$. Then one can mimic the argument following (E.137) to conclude (E.132).

The results in (E.135)-(E.136) follow by similar arguments, with proper redefinition of τ in equation (E.138). □

Lemma E.7: *Let Assumptions 4.3 and 4.5 hold. Let (P_n, θ_n) have the almost sure representations given in Lemma E.1, let $\mathcal{J}^*$ be defined as in (E.29), and assume that $\mathcal{J}^* \neq \emptyset$. Let $\widetilde{\mathcal{C}}$ collect all size d subsets C of $\{1, ..., J+2d+2\}$ ordered lexicographically by their smallest, then second smallest, etc. elements. Let the random variable $\mathcal{C}$ equal the first element of $\widetilde{\mathcal{C}}$ s.t. $\det K^C \neq 0$ and $\lambda^C = (K^C)^{-1}g^C \in \mathfrak{W}^{*,-\delta}(0)$ if such an element exists; else, let $\mathcal{C} = \{J+1, ..., J+d\}$ and $\lambda^C = \mathbf{1}_d$, where $\mathbf{1}_d$ denotes a d vector with each entry equal to 1, and K, g and $\mathfrak{W}^{*,-\delta}$ are as defined in Lemma E.2. Then, for any $\eta > 0$, there exist $0 < \varepsilon_\eta < \infty$ and $N \in \mathbb{N}$ s.t. $n \geq N$ implies*

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \eta. \tag{E.152}$$

Proof. We bound the probability in (E.152) as follows:

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0)\neq\emptyset, \left|\det K^{\mathcal{C}}\right|\leq\varepsilon_\eta\right) \leq \mathbf{P}\left(\exists C\in\widetilde{\mathcal{C}}:\lambda^C\in B^d, \left|\det K^C\right|\leq\varepsilon_\eta\right) \tag{E.153}$$

$$\leq \sum_{C\in\widetilde{\mathcal{C}}:\left|\det K^C\right|\leq\varepsilon_\eta} \mathbf{P}\left(\lambda^C\in B^d\right) \tag{E.154}$$

$$\leq \sum_{C\in\widetilde{\mathcal{C}}:|\alpha^C|\leq\varepsilon_\eta^{2/d}} \mathbf{P}\left(\lambda^C\in B^d\right), \tag{E.155}$$

where α^C denote the smallest eigenvalue of $K^C K^{C\prime}$. Here, the first inequality holds because $\mathfrak{W}^{*,-\delta}\subseteq B^d$ and so the event in the first probability implies the event in the next one; the second inequality is Boolean algebra; the last inequality follows because $|\det K^C|\geq|\alpha^C|^{d/2}$. Noting that $\widetilde{\mathcal{C}}$ has $\binom{J+2d+2}{d}$ elements, it suffices to show that

$$\left|\alpha^C\right|\leq\varepsilon_\eta^{2/d}\Longrightarrow\mathbf{P}\left(\lambda^C\in B^d\right)\leq\overline{\eta}\equiv\frac{\eta}{\binom{J+2d+2}{d}}.$$

Thus, fix $C\in\widetilde{\mathcal{C}}$. Let q^C denote the eigenvector associated with α^C and recall that because $K^C K^{C\prime}$ is symmetric, $\left\|q^C\right\|=1$. Thus the claim is equivalent to:

$$|q^{C\prime}K^C K^{C\prime}q^C|\leq\varepsilon_\eta^{2/d} \implies \mathbf{P}((K^C)^{-1}g^C\in\mathfrak{B}^d_\rho)\leq\overline{\eta}. \tag{E.156}$$

Now, if $|q^{C\prime}K^C K^{C\prime}q^C|\leq\varepsilon_\eta^{2/d}$ and $(K^C)^{-1}g^C\in\mathfrak{B}^d_\rho$, then the Cauchy-Schwarz inequality yields

$$\left|q^{C\prime}g^C_{P_n}\right|=\left|q^{C\prime}K^C\left(K^C\right)^{-1}g^C\right|<\sqrt{d}\varepsilon_\eta^{1/d}, \tag{E.157}$$

hence

$$\mathbf{P}((K^C)^{-1}g^C\in\mathfrak{B}^d_\rho)\leq\mathbf{P}\Big(|q^{C\prime}g^C|<\sqrt{d}\varepsilon_\eta^{1/d}\Big). \tag{E.158}$$

If q^C assigns non-zero weight only to non-stochastic constraints, the result follows immediately. If q^C assigns non-zero weight also to stochastic constraints, Assumptions 4.3 and 4.5 (iii) yield

$$\begin{aligned}
&\mathrm{eig}(\tilde{\Omega})\geq\omega\\
&\Longrightarrow Var_{\mathbf{P}}(q^{C\prime}g^C)\geq\omega\\
&\Longrightarrow \mathbf{P}\left(|q^{C\prime}g^C|<\sqrt{d}\varepsilon_\eta^{1/d}\right)=\mathbf{P}\left(-\sqrt{d}\varepsilon_\eta^{1/d}<q^{C\prime}g^C<\sqrt{d}\varepsilon_\eta^{1/d}\right)\\
&\qquad<\frac{2\sqrt{d}\varepsilon_\eta^{1/d}}{\sqrt{2\omega\pi}},
\end{aligned} \tag{E.159}$$

where the result in (E.159) uses that the density of a normal r.v. is maximized at the expected value. The result follows by choosing

$$\varepsilon_\eta=\left(\frac{\overline{\eta}\sqrt{2\omega\pi}}{2\sqrt{d}}\right)^d.$$

□

LEMMA E.8: *Let Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. If $J_2\geq d$, then $\exists\underline{c}>0$ s.t.*

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}}\inf_{\theta\in\Theta_I(P)}P(c_n^I(\theta)\geq\underline{c})=1.$$

Proof. Fix any $c\geq0$ and restrict attention to constraints $\{J_1+1,...,J_1+d,J_1+J_2+1,...,J_1+J_2+d\}$, i.e.

the inequalities that jointly correspond to the first d equalities. We separately analyze the case when (i) the corresponding estimated gradients $\{\hat{D}_{n,j}(\theta) : j = J_1+1,...,J_1+d\}$ are linearly independent and (ii) they are not. If $\{\hat{D}_{n,j}(\theta) : j = J_1+1,...,J_1+d\}$ converge to linearly independent limits, then only the former case occurs infinitely often; else, both may occur infinitely often, and we conduct the argument along two separate subsequences if necessary.

For the remainder of this proof, because the sequence $\{\theta_n\}$ is fixed and plays no direct role in the proof, we suppress dependence of $\hat{D}_{n,j}(\theta)$ and $\mathbb{G}^b_{n,j}(\theta)$ on θ. Also, if C is an index set picking certain constraints, then $\hat{D}^C_n$ is the matrix collecting the corresponding estimated gradients, and similarly for $\mathbb{G}^{b,C}_n$.

Suppose now case (i), then there exists an index set $\bar{C} \subset \{J_1+1,...,J_1+d, J_1+J_2+1,\dots,J_1+J_2+d\}$ picking one direction of each constraint s.t. p is a positive linear combination of the rows of $\hat{D}^{\bar{C}}_P$. (This choice ensures that a Karush-Kuhn-Tucker condition holds, justifying the step from (E.160) to (E.161) below.) Then the coverage probability $P^*(V^I_n(\theta,c) \neq \emptyset)$ is asymptotically bounded above by

$$P^*\Big(\sup_{\lambda\in\rho B^d_{n,\rho}} \Big\{p'\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\mathcal{J}^*\Big\} \ge 0\Big) \le P^*\Big(\sup_{\lambda\in\mathbb{R}^d} \Big\{p'\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\Big\} \ge 0\Big) \tag{E.160}$$

$$= P^*\Big(p'(\hat{D}^{\bar{C}}_n)^{-1}(c\mathbf{1}_d - \mathbb{G}^{b,\bar{C}}_n) \ge 0\Big) \tag{E.161}$$

$$= P^*\left(\frac{p'(\hat{D}^{\bar{C}}_n)^{-1}(c\mathbf{1}_d - \mathbb{G}^{b,\bar{C}}_n)}{\sqrt{p'(\hat{D}^{\bar{C}}_n)^{-1}\Omega^C_P(\hat{D}^{\bar{C}}_n)^{-1}p}} \ge 0\right) \tag{E.162}$$

$$= P^*\left(\frac{p'adj(\hat{D}^{\bar{C}}_n)(c\mathbf{1}_d - \mathbb{G}^{b,\bar{C}}_n)}{\sqrt{p'(adj(\hat{D}^{\bar{C}}_n)\Omega^C_P adj(\hat{D}^{\bar{C}}_n)p}} \ge 0\right) \tag{E.163}$$

$$= \Phi\left(\frac{p'adj(\hat{D}^{\bar{C}}_n)c\mathbf{1}_d}{\sqrt{p'(adj(\hat{D}^{\bar{C}}_n)\Omega^C_P adj(\hat{D}^{\bar{C}}_n)p}}\right) + o_{\mathcal{P}}(1) \tag{E.164}$$

$$\le \Phi(d\omega^{-1/2}c) + o_{\mathcal{P}}(1). \tag{E.165}$$

Here, (E.160) removes constraints and hence enlarges the feasible set; (E.161) solves in closed form; (E.162) divides through by a positive scalar; (E.163) eliminates the determinant of $\hat{D}^{\bar{C}}_n$, using that rows of $\hat{D}^{\bar{C}}_n$ can always be rearranged so that the determinant is positive; (E.164) follows by Assumption 4.5, using that the term multiplying $\mathbb{G}^{b,\bar{C}}_n$ is $O_{\mathcal{P}}(1)$; and (E.165) uses that by Assumption 4.3, there exists a constant $\omega > 0$ that does not depend on θ such that the smallest eigenvalue of Ω_P is bounded from below by ω. The result follows for any choice of $\underline{c} \in (0, \Phi^{-1}(1-\alpha)\times\omega^{1/2}/d)$.

In case (ii), there exists an index set $\bar{C} \subset \{J_1+2,...,J_1+d, J_1+J_2+2,...,J_1+J_2+d\}$ collecting $d-1$ or fewer linearly independent constraints s.t. $\hat{D}_{n,J_1+1}$ is a positive linear combination of the rows of $\hat{D}^{\bar{C}}_P$. (Note that $\bar{C}$ cannot contain J_1+1 or J_1+J_2+1.) One can then write

$$P^*\Big(\sup_{\lambda\in\rho B^d_{n,\rho}} \Big\{p'\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\cup\{J_1+J_2+1\}\Big\} \ge 0\Big) \tag{E.166}$$

$$\le P^*\Big(\exists\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\cup\{J_1+J_2+1\}\Big) \tag{E.167}$$

$$\le P^*\Big(\sup_{\lambda\in\rho B^d_{n,\rho}} \Big\{\hat{D}_{n,J_1+1}\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\Big\} \ge \inf_{\lambda\in\rho B^d_{n,\rho}} \Big\{\hat{D}_{n,J_1+1}\lambda : \hat{D}_{n,J_1+J_2+1}\lambda \le c - \mathbb{G}^b_{n,J_1+J_2+1}\Big\}\Big) \tag{E.168}$$

$$= P^*\Big(\hat{D}_{n,J_1+1}\hat{D}^{\bar{C}\prime}_n(\hat{D}^{\bar{C}}_n\hat{D}^{\bar{C}\prime}_n)^{-1}(c\mathbf{1}_{\bar{d}} - \mathbb{G}^{b,\bar{C}}_n) \ge -c + \mathbb{G}^b_{n,J_1+J_2+1}\Big). \tag{E.169}$$

Here, the reasoning from (E.166) to (E.168) holds because we evaluate the probability of increasingly larger events; in particular, if the event in (E.168) fails, then the constraint sets corresponding to the sup and inf can be separated by a hyperplane with gradient $\hat{D}_{n,J_1+1}$ and so cannot intersect. The last step solves the optimization problems in closed form, using (for the sup) that a Karush-Kuhn-Tucker condition again holds by construction and (for the inf) that $\hat{D}_{n,J_1+J_2+1} = -\hat{D}_{n,J_1+1}$. Expression (E.169) resembles (E.162), and the argument can be concluded in analogy to (E.163)-(E.165). □

LEMMA E.9: *Let Assumptions 4.1, 4.2, 4.3-(II), 4.4, and 4.5 hold. Suppose that both $\pi_{1,j}$ and $\pi_{1,j+R_1}$ are finite, with $\pi_{1,j}$, $j = 1,\dots,J$, defined in (D.4). Let (P_n, θ_n) be the sequence satisfying the conditions of Lemma E.3. Then for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$,*

(1) $\sigma^2_{P_n,j}(\theta'_n)/\sigma^2_{P_n,j+R_1}(\theta'_n) \to 1$ *for* $j = 1,\cdots,R_1$.

(2) $Corr_{P_n}(m_j(X_i,\theta'_n), m_{j+R_1}(X_i,\theta'_n)) \to -1$ *for* $j = 1,\cdots,R_1$.

(3) $|\mathbb{G}_{n,j}(\theta'_n) + \mathbb{G}_{n,j+R_1}(\theta'_n)| \xrightarrow{P_n} 0$, *and* $|\mathbb{G}^b_{n,j}(\theta'_n) + \mathbb{G}^b_{n,j+R_1}(\theta'_n)| \xrightarrow{P^*_n} 0$ *for almost all* $\{X_i\}_{i=1}^\infty$.

(4) $\rho\|D_{P_n,j+R_1}(\theta'_n) + D_{P_n,j}(\theta'_n)\| \to 0$.

Proof. By Lemma E.5, for each j, $\lim_{n\to\infty} \kappa_n^{-1} \frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\sigma_{P_n,j}(\theta'_n)} = \pi_{1,j}$, and hence the condition that $\pi_{1,j}, \pi_{1,j+R_1}$ are finite is inherited by the limit of the corresponding sequences $\kappa_n^{-1} \frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta'_n)]}{\sigma_{P_n,j}(\theta'_n)}$ and $\kappa_n^{-1} \frac{\sqrt{n}E_{P_n}[m_{j+J11}(X_i,\theta'_n)]}{\sigma_{P_n,j+J11}(\theta'_n)}$.

We first establish Claims 1 and 2. We consider two cases.

Case 1.

$$\lim_{n\to\infty} \frac{\kappa_n}{\sqrt{n}} \sigma_{P_n,j}(\theta'_n) > 0, \tag{E.170}$$

which implies that $\sigma_{P_n,j}(\theta'_n) \to \infty$ at rate $\sqrt{n}/\kappa_n$ or faster. Claim 1 then holds because

$$\frac{\sigma^2_{P_n,j+R_1}(\theta'_n)}{\sigma^2_{P_n,j}(\theta'_n)} = \frac{\sigma^2_{P_n,j}(\theta'_n) + Var_{P_n}(t_j(X_i,\theta'_n)) + 2Cov_{P_n}(m_j(X_i,\theta'_n), t_j(X_i,\theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)} \to 1, \tag{E.171}$$

where the convergence follows because $Var_{P_n}(t_j(X_i,\theta'_n))$ is bounded due to Assumption 4.3-(II),

$$|Cov_{P_n}(m_j(X_i,\theta'_n), t_j(X_i,\theta'_n))/\sigma^2_{P_n,j}(\theta'_n)| \le (Var_{P_n}(t_j(X_i,\theta'_n)))^{1/2}/\sigma_{P_n,j}(\theta'_n),$$

and the fact that $\sigma_{P_n,j}(\theta'_n) \to \infty$. A similar argument yields Claim 2.

Case 2.

$$\lim_{n\to\infty} \frac{\kappa_n}{\sqrt{n}} \sigma_{P_n,j}(\theta'_n) = 0. \tag{E.172}$$

In this case, $\pi_{1,j}$ being finite implies that $E_{P_n} m_j(X_i,\theta'_n) \to 0$. Again using the upper bound on $t_j(X_i,\theta'_n)$ similarly to (E.171), it also follows that

$$\lim_{n\to\infty} \frac{\kappa_n}{\sqrt{n}} \sigma_{P_n,j+R_1}(\theta'_n) = 0, \tag{E.173}$$

and hence that $E_{P_n}(t_j(X_i,\theta'_n)) \to 0$. We then have, using Assumption 4.3-(II) again,

$$\begin{aligned} Var_{P_n}(t_j(X_i,\theta'_n)) &= \int t_j(x,\theta'_n)^2 dP_n(x) - E_{P_n}[t_j(X_i,\theta'_n)]^2 \\ &\le M \int t_j(x,\theta'_n) dP_n(x) - E_{P_n}[t_j(X_i,\theta'_n)]^2 \to 0. \end{aligned} \tag{E.174}$$

Hence,

$$\frac{\sigma^2_{P_n,j+R_1}(\theta'_n)}{\sigma^2_{P_n,j}(\theta'_n)} = \frac{\sigma^2_{P_n,j}(\theta'_n) + Var_{P_n}(t_j(X_i,\theta'_n)) + 2Cov_{P_n}(m_j(X_i,\theta'_n), t_j(X_i,\theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)}$$
$$\leq \frac{\sigma^2_{P_n,j}(\theta'_n) + Var_{P_n}(t_j(X_i,\theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)} + \frac{2(Var_{P_n}(t_j(X_i,\theta'_n)))^{1/2}}{\sigma_{P_n,j}(\theta'_n)}$$
$$\to 1, \tag{E.175}$$

and the first claim follows.

To obtain claim 2, note that

$$Corr_{P_n}(m_j(X_i,\theta'_n), m_{j+R_1}(X_i,\theta'_n)) = \frac{-\sigma^2_{P_n,j}(\theta'_n) - Cov_{P_n}(m_j(X_i,\theta'_n), t_j(X_i,\theta'_n))}{\sigma_{P_n,j}(\theta'_n)\sigma_{P_n,j+R_1}(\theta'_n)}$$
$$\to -1, \tag{E.176}$$

where the result follows from (E.174) and (E.175).

To establish Claim 3, consider $\mathbb{G}_n$ below. Note that, for $j = 1, \cdots, R_1$,

$$\begin{bmatrix} \mathbb{G}_{n,j}(\theta'_n) \\ \mathbb{G}_{n,j+R_1}(\theta'_n) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (m_j(X_i,\theta'_n) - E_{P_n}[m_j(X_i,\theta'_n)])}{\sigma_{P_n,j}(\theta'_n)} \\ -\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (m_j(X_i,\theta'_n) - E_{P_n}[m_j(X_i,\theta'_n)]) + \frac{1}{\sqrt{n}}\sum_{i=1}^n (t_j(X_i,\theta'_n) - E_{P_n}[t_j(X_i,\theta'_n)])}{\sigma_{P_n,j+R_1}(\theta'_n)} \end{bmatrix}. \tag{E.177}$$

Under the conditions of Case 1 above, we immediately obtain

$$|\mathbb{G}_{n,j}(\theta'_n) + \mathbb{G}_{n,j+R_1}(\theta'_n)| \xrightarrow{P_n} 0. \tag{E.178}$$

Under the conditions in Case 2 above, $\frac{1}{\sqrt{n}}\sum_{i=1}^n (t_j(X_i,\theta'_n) - E_{P_n}[t_j(X_i,\theta'_n)] = o_{\mathcal{P}}(1)$ due to the variance of this term being equal to $Var_{P_n}(t_j(X_i,\theta'_n)) \to 0$ and Chebyshev's inequality. Therefore, (E.178) obtains again. These results imply that $\mathbb{Z}_j + \mathbb{Z}_{j+R_1} = 0, a.s.$ By Lemma E.15, $\{\mathbb{G}_n^b\}$ converges in law to the same limit as $\{\mathbb{G}_n\}$ for almost all sample paths $\{X_i\}_{i=1}^\infty$. This and (E.178) then imply the second half of Claim 3.

To establish Claim 4, finiteness of $\pi_{1,j}$ and $\pi_{1,j+R_1}$ implies that

$$E_{P_n}\left(\frac{m_j(X,\theta'_n)}{\sigma_{P_n,j}(\theta'_n)} + \frac{m_{j+R_1}(X,\theta'_n)}{\sigma_{P_n,j+R_1}(\theta'_n)}\right) = O_{\mathcal{P}}\left(\frac{\kappa_n}{\sqrt{n}}\right). \tag{E.179}$$

Define the $1 \times d$ vector

$$q_n \equiv D_{P_n,j+R_1}(\theta'_n) + D_{P_n,j}(\theta'_n). \tag{E.180}$$

Suppose by contradiction that

$$\rho q_n \to \varsigma \neq 0,$$

where $\|\varsigma\|$ might be infinite. Write

$$\tilde{r}_n = \frac{q'_n}{\|q_n\|}. \tag{E.181}$$

Let

$$r_n = \tilde{r}_n \rho \kappa_n^2/\sqrt{n}. \tag{E.182}$$

Using a mean value expansion (where $\bar{\theta}_n$ and $\tilde{\theta}_n$ in the expressions below are two potentially different vectors that lie component-wise between θ_n' and $\theta_n' + r_n$) we obtain

$$E_{P_n}\left(\frac{m_j(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}+\frac{m_{j+R_1}(X,\theta_n'+r_n)}{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}\right)=E_{P_n}\left(\frac{m_j(X,\theta_n')}{\sigma_{P_n,j}(\theta_n')}+\frac{m_{j+R_1}(X,\theta_n')}{\sigma_{P_n,j+R_1}(\theta_n')}\right)+\Big(D_{P_n,j}(\bar{\theta}_n)+D_{P_n,j+R_1}(\tilde{\theta}_n)\Big)r_n$$

$$=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+(D_{P_n,j}(\theta_n')+D_{P_n,j+R_1}(\theta_n'))r_n+\big(D_{P_n,j}(\bar{\theta}_n)-D_{P_n,j}(\theta_n')\big)r_n+\Big(D_{P_n,j+R_1}(\tilde{\theta}_n)-D_{P_n,j+R_1}(\theta_n')\Big)r_n$$

$$=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+\frac{\rho\kappa_n^2}{\sqrt{n}}+O_{\mathcal{P}}(\frac{\rho^2\kappa_n^4}{n}). \tag{E.183}$$

It then follows that there exists $N \in \mathbb{N}$ such that for all $n \geq N$, the right hand side in (E.183) is strictly greater than zero.

Next, observe that

$$\begin{aligned}
&E_{P_n}\left(\frac{m_j(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}+\frac{m_{j+R_1}(X,\theta_n'+r_n)}{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}\right)\\
=&E_{P_n}\left(\frac{m_j(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}+\frac{m_{j+R_1}(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}\right)-\left(\frac{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}-1\right)\frac{E_{P_n}(m_{j+R_1}(X,\theta_n'+r_n))}{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}\\
=&E_{P_n}\left(\frac{m_j(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}+\frac{m_{j+R_1}(X,\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}\right)-o_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}}).
\end{aligned} \tag{E.184}$$

Here, the last step is established as follows. First, using that $\sigma_{P_n,j}(\theta_n'+r_n)$ is bounded away from zero for n large enough by the continuity of $\sigma(\cdot)$ and Assumption 4.3-(II), we have

$$\frac{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}{\sigma_{P_n,j}(\theta_n'+r_n)}-1=\frac{\sigma_{P_n,j+R_1}(\theta_n')}{\sigma_{P_n,j}(\theta_n')}-1+o_{\mathcal{P}}(1)=o_{\mathcal{P}}(1), \tag{E.185}$$

where we used Claim 1. Second, using Assumption 4.4, we have that

$$\frac{E_{P_n}(m_{j+R_1}(X,\theta_n'+r_n))}{\sigma_{P_n,j+R_1}(\theta_n'+r_n)}=\frac{E_{P_n}(m_{j+R_1}(X,\theta_n'))}{\sigma_{P_n,j+R_1}(\theta_n')}+D_{P_n,j+R_1}(\tilde{\theta}_n)r_n=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+O_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}}). \tag{E.186}$$

The product of (E.185) and (E.186) is therefore $o_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}})$ and (E.184) follows.

To conclude the argument, note that for n large enough, $m_{j+R_1}(X,\theta_n'+r_n) \leq -m_j(X,\theta_n'+r_n)$ *a.s.* because for any $\theta_n \in \Theta_I(P_n)$ and $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$ for n large enough, $\theta_n' + r_n \in \Theta^\epsilon$ and Assumption 4.3-(II) applies. Therefore, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, the left hand side in (E.183) is strictly less than the right hand side, yielding a contradiction. □

Below, we let $\mathcal{R}_1 = \{1, \cdots, R_1\}$ and $\mathcal{R}_2 = \{R_1+1, \cdots, 2R_1\}$.

Lemma E.10: *Suppose Assumptions 4.1, 4.2, and 4.5 hold. For each $\theta \in \Theta$, let $\eta_{n,j}(\theta) = \sigma_{P,j}(\theta)/\hat{\sigma}_{n,j}(\theta) - 1$. Then, (i) for each $j = 1, \ldots, J_1 + J_2$*

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \to 0\Big) = 1. \tag{E.187}$$

(ii) For any $j = 1, \ldots, R_1$ let

$$\hat{\sigma}^M_{n,j}(\theta) = \hat{\sigma}^M_{n,j+R_1}(\theta) \equiv \hat{\mu}_{n,j}(\theta)\hat{\sigma}_{n,j}(\theta) + (1-\hat{\mu}_{n,j}(\theta))\hat{\sigma}_{n,j+R_1}(\theta). \tag{E.188}$$

Let (P_n, θ_n) be a sequence such that $P_n \in \mathcal{P}$, $\theta_n \in \Theta$ for all n, and $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}$. Let $\mathcal{J}^*$ be *defined as in* (E.29)*. Then, for any* $\eta > 0$*, there exists* $N \in \mathbb{N}$ *such that*

$$P_n\Big(\max_{j\in(\mathcal{R}_1\cup\mathcal{R}_2)\cap\mathcal{J}^*}\Big|\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}-1\Big|>\eta\Big)<\eta \tag{E.189}$$

for all $n \geq N$*.*

Proof. We first show that, for any $\epsilon > 0$ and for any $j = 1, \ldots, J_1 + J_2$,

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}-1\Big|\leq\epsilon\Big)\to 1. \tag{E.190}$$

For this, define the following sets:

$$\mathcal{M}_j \equiv \{m_j(\cdot,\theta)/\sigma_{P,j}(\theta) : \theta\in\Theta, P\in\mathcal{P}\} \tag{E.191}$$

$$\mathcal{S}_j \equiv \{(m_j(\cdot,\theta)/\sigma_{P,j}(\theta))^2 : \theta\in\Theta, P\in\mathcal{P}\}. \tag{E.192}$$

By Assumptions 4.1-(a), 4.1 (iv), 4.5 (i), (iii), and arguing as in the proof of Lemma D.2.2 (and D.2.1) in Bugni, Canay, and Shi (2015), it follows that $\mathcal{S}_j$ and $\mathcal{M}_j$ are Glivenko-Cantelli (GC) classes uniformly in $P \in \mathcal{P}$ (in the sense of van der Vaart and Wellner, 2000, page 167).

Therefore, for any $\epsilon > 0$,

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma^2_{P,j}(\theta)}-\frac{E_P[m_j(X,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big|\leq\epsilon\Big)\to 1 \tag{E.193}$$

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta)-E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big|\leq\epsilon\Big)\to 1. \tag{E.194}$$

Note that, by Assumption 4.1 (iv), $|E_P[m_j(X,\theta)]/\sigma_{P,j}(\theta)| \leq M$ for some constant $M > 0$ that does not depend on P and $(x^2 - y^2) \leq |x+y||x-y| \leq 2M|x-y|$ for all $x, y \in [-M, M]$. By (E.194), for any $\epsilon > 0$, it follows that

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta)^2-E_P[m_j(X,\theta)]^2}{\sigma^2_{P,j}(\theta)}\Big|\leq\epsilon\Big)\to 1. \tag{E.195}$$

By the uniform continuity of $x \mapsto \sqrt{x}$ on $\mathbb{R}_+$, for any $\epsilon > 0$, there is a constant $\eta > 0$ such that

$$\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)}-1\Big|\leq\eta \ \Rightarrow \Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}-1\Big|\leq\epsilon. \tag{E.196}$$

By the definition of $\sigma^2_{P,j}(\theta)$ and the triangle inequality,

$$\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)}-1\Big|\leq\Big|\frac{n^{-1}\sum_{i=1}^n m(X_i,\theta)^2-E[m_j(X_i,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big|+\Big|\frac{\bar{m}_{n,j}(\theta)^2-E[m_j(X_i,\theta)]^2}{\sigma^2_{P,j}(\theta)}\Big|. \tag{E.197}$$

By (E.196)-(E.197), bounding each of the terms on the right hand side of (E.197) by $\eta/2$ implies $|\hat{\sigma}_{n,j}(\theta)/\sigma_{P,j}(\theta) - 1| \leq \epsilon$. This, together with (E.193) and (E.195), ensures that, for any $\epsilon > 0$, (E.190) holds.

Note that $|\hat{\sigma}_{n,j}(\theta)/\sigma_{P,j}(\theta) - 1| \leq \epsilon$ implies $\hat{\sigma}_{n,j}(\theta) > 0$, and argue as in the proof of Lemma D.2.4 in Bugni, Canay, and Shi (2015) to conclude that

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\sigma_{P,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}-1\Big|\leq\epsilon\Big)\to 1. \tag{E.198}$$

Finally, recall that $\eta_{n,j}(\theta) = \sigma_{P,j}(\theta)/\hat{\sigma}_{n,j}(\theta) - 1$ and note that for any $\epsilon > 0$,

$$\begin{aligned}
1 &= \lim_{n\to\infty} \inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| \leq \epsilon\Big) \\
&\leq \inf_{P\in\mathcal{P}} \lim_{n\to\infty} P\Big(\bigcap_{m\geq n}\big\{\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| \leq \epsilon\big\}\Big) \\
&= \inf_{P\in\mathcal{P}} P\Big(\lim_{n\to\infty} \bigcap_{m\geq n}\{\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| \leq \epsilon\}\Big) \\
&= \inf_{P\in\mathcal{P}} P\Big(\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| \leq \epsilon, \text{ for almost all } n\Big), \qquad \text{(E.199)}
\end{aligned}$$

where the second equality is due to the continuity of probability with respect to monotone sequences. Therefore, the first conclusion of the lemma follows.

(ii) We first give the limit of $\hat{\mu}_{n,j}(\theta_n)$. Recall the definitions of $\hat{\mu}_{n,j+R_1}$ and $\hat{\mu}_{n,j}(\theta_n)$ in (E.14)-(E.15). Note that

$$\begin{aligned}
&\sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d} \Big|\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n')}{\hat{\sigma}_{n,j}(\theta_n')} - \kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\Big| \\
&\leq \sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d} \Big|\kappa_n^{-1}\frac{\sqrt{n}(\bar{m}_{n,j}(\theta_n') - E_{P_n}[m_j(X_i,\theta_n')])}{\sigma_{n,j}(\theta_n')}(1+\eta_{n,j}(\theta_n')) + \kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\Big| \\
&\leq \sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d} |\kappa_n^{-1}\mathbb{G}_n(\theta_n')(1+\eta_{n,j}(\theta_n'))| + \Big|\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\kappa_n\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\Big| = o_{\mathcal{P}}(1), \qquad \text{(E.200)}
\end{aligned}$$

where the last equality follows from $\sup_{\theta\in\Theta}|\mathbb{G}_n(\theta)| = O_{\mathcal{P}}(1)$ due to asymptotic tightness of $\{\mathbb{G}_n\}$ (uniformly in P) by Lemma D.1 in Bugni, Canay, and Shi (2015), Theorem 3.6.1 and Lemma 1.3.8 in van der Vaart and Wellner (2000), and $\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| = o_{\mathcal{P}}(1)$ by part (i) of this Lemma. Hence,

$$\hat{\mu}_{n,j}(\theta_n) \stackrel{P_n}{\to} 1 - \min\Big\{\max(0, \frac{\pi_{1,j}}{\pi_{1,j+R_1}+\pi_{1,j}}), 1\Big\}, \qquad \text{(E.201)}$$

unless $\pi_{1,j+R_1} + \pi_{1,j} = 0$ (this case is considered later). This implies that if $\pi_{1,j} \in (-\infty, 0]$ and $\pi_{1,j+R_1} = -\infty$, one has

$$\hat{\mu}_{n,j}(\theta_n) \stackrel{P_n}{\to} 1. \qquad \text{(E.202)}$$

Similarly, if $\pi_{1,j} = -\infty$ and $\pi_{1,j+R_1} \in (-\infty, 0]$, one has

$$\hat{\mu}_{n,j+R_1}(\theta_n) \stackrel{P_n}{\to} 1. \qquad \text{(E.203)}$$

Now, one may write

$$\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1 = \frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)}\Big(\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1\Big) + \Big(\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)} - 1\Big) = O_{P_n}(1)\Big(\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1\Big) + o_{P_n}(1), \qquad \text{(E.204)}$$

where the second equality follows from the first conclusion of the lemma. Hence, for the second conclusion of the lemma, it suffices to show $\hat{\sigma}_{n,j}(\theta_n)/\hat{\sigma}^M_{n,j}(\theta_n) - 1 = o_{\mathcal{P}}(1)$. For this, we consider three cases.

Suppose first $j \in \mathcal{R}_1 \cap \mathcal{J}^*$ and $j + R_1 \notin \mathcal{J}^*$. Then, $\pi^*_{1,j} = 0$ and $\pi^*_{1,j+R_1} = -\infty$. Then,

$$\hat{\sigma}^M_{n,j}(\theta_n) = \hat{\mu}_{n,j}(\theta_n)\hat{\sigma}_{n,j}(\theta_n) + (1 - \hat{\mu}_{n,j}(\theta_n))\hat{\sigma}_{n,j+R_1}(\theta_n) \tag{E.205}$$

$$= (1 + o_{P_n}(1))\hat{\sigma}_{n,j}(\theta_n) + (1 - \hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)), \tag{E.206}$$

where the second equality follows from (E.202) and the fact that

$$\hat{\sigma}_{n,j+R_1}(\theta_n) = \Big(\hat{\sigma}^2_{n,j}(\theta_n) + 2\widehat{Cov}_n(m_j(X_i,\theta), t_j(X_i,\theta)) + \widehat{Var}_n(t_j(X_i,\theta))\Big)^{1/2}$$

$$= \Big(\hat{\sigma}^2_{n,j}(\theta_n) + O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)) + O_{P_n}(1)\Big)^{1/2} = O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)), \tag{E.207}$$

where the second equality follows from, $Var_{P_n}(t_j(X_i,\theta))$ being bounded by Assumption 4.3-(II) and

$$\widehat{Var}_n(t_j(X_i,\theta)) = Var_{P_n}(t_j(X_i,\theta)) + o_{P_n}(1) \tag{E.208}$$

$$\widehat{Cov}_n(m_j(X_i,\theta), t_j(X_i,\theta)) \le \hat{\sigma}_{n,j}(\theta_n)\widehat{Var}_n(t_j(X_i,\theta))^{1/2}, \tag{E.209}$$

where the last inequality is due to the Cauchy-Schwarz inequality.

Therefore,

$$\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1 = \frac{\hat{\sigma}_{n,j}(\theta_n) - \hat{\sigma}^M_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} = \frac{(1 - \hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n))}{(1 + o_{P_n}(1))\hat{\sigma}_{n,j}(\theta_n) + (1 - \hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n))} = o_{P_n}(1), \tag{E.210}$$

where we used $\hat{\sigma}^{-1}_{n,j}(\theta_n) = O_{P_n}(1)$ by equation (4.3) and part (i) of the lemma. By (E.204) and (E.210), $\sigma_{P_n,j}(\theta_n)/\hat{\sigma}^M_{n,j}(\theta_n) - 1 = o_{P_n}(1)$. Using a similar argument, the same conclusion follows when $j \in \mathcal{R}_1, j \notin \mathcal{J}^*$, but $j + R_1 \in \mathcal{R}_2 \cap \mathcal{J}^*$.

Now consider the case $j \in \mathcal{R}_1 \cap \mathcal{J}^*$ and $j + R_1 \in \mathcal{R}_2 \cap \mathcal{J}^*$. Then, $\pi^*_{1,j} = 0$ and $\pi^*_{1,j+R_1} = 0$. In this case, $\hat{\mu}_{n,j}(\theta_n) \in [0,1]$ for all n and by Lemma E.9 (1),

$$\Big|\frac{\sigma_{P_n,j}(\theta_n)}{\sigma_{P_n,j+R_1}(\theta_n)} - 1\Big| = o_{P_n}(1), \quad \text{for } j = 1, \cdots, R_1, \tag{E.211}$$

and therefore,

$$\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1 = \frac{\sigma_{P_n,j}(\theta_n) - \hat{\sigma}^M_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}$$

$$= \frac{[\hat{\mu}_{n,j}(\theta_n) + (1 - \hat{\mu}_{n,j}(\theta_n))]\sigma_{P_n,j}(\theta_n) - [\hat{\mu}_{n,j}(\theta_n)\hat{\sigma}_{n,j}(\theta_n) + (1 - \hat{\mu}_{n,j}(\theta_n))\hat{\sigma}_{n,j+R_1}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)}$$

$$= \frac{\hat{\mu}_{n,j}(\theta_n)[\sigma_{P_n,j}(\theta_n) - \hat{\sigma}_{n,j}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)} + \frac{(1 - \hat{\mu}_{n,j}(\theta_n))[\sigma_{P_n,j+R_1}(\theta_n) - \hat{\sigma}_{n,j+R_1}(\theta_n) + o_{P_n}(1)]}{\hat{\sigma}^M_{n,j}(\theta_n)}, \tag{E.212}$$

where the second equality follows from the definition of $\hat{\sigma}^M_{n,j}(\theta_n)$, and the third equality follows from (E.211) and $\sigma_{P_n,j+R_1}$ bounded away from 0 due to (4.3). Note that

$$\frac{\hat{\mu}_{n,j}(\theta_n)[\sigma_{P_n,j}(\theta_n) - \hat{\sigma}_{n,j}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)} = \hat{\mu}_{n,j}(\theta_n)\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}\Big(\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)} - 1\Big) = o_{P_n}(1), \tag{E.213}$$

where the second equality follows from the first conclusion of the lemma. Similarly,

$$\frac{(1-\hat{\mu}_{n,j}(\theta_n))[\sigma_{P_n,j+R_1}(\theta_n)-\hat{\sigma}_{n,j+R_1}(\theta_n)+o_{P_n}(1)]}{\hat{\sigma}^M_{n,j}(\theta_n)}$$
$$=(1-\hat{\mu}_{n,j}(\theta_n))\frac{\hat{\sigma}_{n,j+R_1}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}\Big(\frac{\sigma_{P_n,j+R_1}(\theta_n)}{\hat{\sigma}_{n,j+R_1}(\theta_n)}-1+o_{P_n}(1)\Big)=o_{P_n}(1). \quad \text{(E.214)}$$

By (E.212)-(E.214), it follows that $\sigma_{P_n,j}(\theta_n)/\hat{\sigma}^M_{n,j}(\theta_n)-1=o_{P_n}(1)$. Therefore, the second conclusion holds for all subcases. □

E.2 Lemmas Used to Prove Theorem B.1

Let $\{X_i^b\}_{i=1}^n$ denote a bootstrap sample drawn randomly from the empirical distribution. Define

$$\mathfrak{G}^b_{n,j}(\theta)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(m_j(X_i^b,\theta)-\bar{m}_n(\theta)\right)/\sigma_{P,j}(\theta)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^n(M_{n,i}-1)m_j(X_i,\theta)/\sigma_{P,j}(\theta), \quad \text{(E.215)}$$

where $\{M_{n,i}\}_{i=1}^n$ denotes the multinomial weights on the original sample, and we let P_n^* denote the conditional distribution of $\{M_{n,i}\}_{i=1}^n$ given the sample path $\{X_i\}_{i=1}^\infty$ (see Appendix E.3 for details on the construction of the bootstrapped empirical process).

Lemma E.11: *(i) Let $\mathcal{M}_P\equiv\{f:\mathcal{X}\to\mathbb{R}:f(\cdot)=\sigma_{P,j}(\theta)^{-1}m_j(\cdot,\theta),\theta\in\Theta,j=1,\cdots,J\}$ and let F be its envelope. Suppose that (i) there exist constants $K,v>0$ that do not depend on P such that*

$$\sup_Q N(\epsilon\|F\|_{L^2_Q},\mathcal{M}_P,L^2_Q)\le K\epsilon^{-v},\ 0<\epsilon<1, \quad \text{(E.216)}$$

where the supremum is taken over all discrete distributions; (ii) There exists a positive constant $\gamma>0$ such that

$$\|(\theta_1,\tilde{\theta}_1)-(\theta_2,\tilde{\theta}_2)\|\le\delta \quad\Rightarrow\quad \sup_{P\in\mathcal{P}}\|Q_P(\theta_1,\tilde{\theta}_1)-Q_P(\theta_2,\tilde{\theta}_2)\|\le M\delta^\gamma. \quad \text{(E.217)}$$

Let δ_n be a positive sequence tending to 0 and let ϵ_n be a positive sequence such that $\epsilon_n/|\delta_n^\gamma\ln\delta_n|\to\infty$ as $n\to\infty$. Then,

$$\sup_{P\in\mathcal{P}}P\left(\sup_{\|\theta-\theta'\|\le\delta_n}\|\mathbb{G}_n(\theta)-\mathbb{G}_n(\theta')\|>\epsilon_n\right)=o(1). \quad \text{(E.218)}$$

Further,

$$\lim_{n\to\infty}P_n^*\left(\sup_{\|\theta-\theta'\|\le\delta_n}\|\mathfrak{G}^b_n(\theta)-\mathfrak{G}^b_n(\theta')\|>\epsilon_n|\{X_i\}_{i=1}^\infty\right)=0. \quad \text{(E.219)}$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$ uniformly in $P\in\mathcal{P}$.

Proof. For the first conclusion of the lemma, it suffices to show that there is a sequence $\{\epsilon_n\}$ such that, uniformly in P:

$$P\left(\sup_{\|\theta-\theta'\|\leq\delta_n}\max_{j=1,\cdots,J}|\mathbb{G}_{n,j}(\theta)-\mathbb{G}_{n,j}(\theta')|>\epsilon_n\right)=o(1). \tag{E.220}$$

For this purpose, we mostly mimic the argument required to show the stochastic equicontinuity of empirical processes (see e.g. van der Vaart and Wellner, 2000, Ch.2.5). Before doing so, note that, arguing as in the proof of Lemma D.1 (Part 1) in Bugni, Canay, and Shi (2015), one has

$$\|\theta-\theta'\|\leq\delta_n \quad\Rightarrow\quad \varrho_P(\theta,\theta')\leq\tilde{\delta}_n, \tag{E.221}$$

where $\tilde{\delta}_n=O(\delta_n^\gamma)$ by assumption. Define

$$\mathcal{M}_{P,\tilde{\delta}_n}=\{\sigma_{P,j}(\theta)^{-1}m_j(\cdot,\theta)-\sigma_{P,j}(\theta')^{-1}m_j(\cdot,\theta')|\theta,\theta'\in\Theta,\varrho_P(\theta,\tilde{\theta})<\tilde{\delta}_n,j=1,\cdots,J\}. \tag{E.222}$$

Define $Z_n(\tilde{\delta}_n)\equiv\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}|\sqrt{n}(\mathbb{P}_n-P)f|$. Then, by (E.221), one has

$$P\left(\sup_{\|\theta-\theta'\|\leq\delta_n}\max_{j=1,\cdots,J}|\mathbb{G}_{n,j}(\theta)-\mathbb{G}_{n,j}(\theta'))|>\epsilon_n)\leq P(Z_n(\tilde{\delta}_n)>\epsilon_n\right). \tag{E.223}$$

From here, we deal with the supremum of empirical processes though symmetrization and an application of a maximal inequality. By Markov's inequality and Lemma 2.3.1 (symmetrization lemma) in van der Vaart and Wellner (2000), one has

$$P(Z_n(\tilde{\delta}_n)>\epsilon_n)\leq\frac{2}{\epsilon_n}E_{P\times P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\right|\right], \tag{E.224}$$

where $\{W_i\}_{i=1}^n$ are i.i.d. Rademacher random variables independent of $\{X_i\}_{i=1}^\infty$ whose law is denoted by P^W. Now, fix the sample path $\{X_i\}_{i=1}^n$, and let $\hat{P}_n$ be the empirical distribution. By Hoeffding's inequality, the stochastic process $f\mapsto\{n^{-1/2}\sum_{i=1}^n W_i f(X_i)\}$ is sub-Gaussian for the $L^2_{\hat{P}_n}$ seminorm $\|f\|_{L^2_{\hat{P}_n}}=(n^{-1}\sum_{i=1}^n f(X_i)^2)^{1/2}$. By the maximal inequality (Corollary 2.2.8) and arguing as in the proof of Theorem 2.5.2 in in van der Vaart and Wellner (2000), one then has

$$\begin{aligned}
E_{P^W}\left[\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\right|\right] &\leq K\int_0^{\tilde{\delta}_n}\sqrt{\ln N(\epsilon,\mathcal{M}_{P,\tilde{\delta}_n},L^2_{\hat{P}_n})}d\epsilon\\
&\leq K\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}\sup_Q\sqrt{\ln N(\epsilon\|F\|_{L^2_Q},\mathcal{M}_P,L^2_Q)}d\epsilon\\
&\leq K'\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}\sqrt{-v\ln\epsilon}d\epsilon,
\end{aligned} \tag{E.225}$$

for some $K'>0$, where the last inequality follows from (E.216). Note that $\sqrt{-\ln\epsilon}\leq-\ln\epsilon$ for $\epsilon\leq\tilde{\delta}_n/\|F\|_{L^2_Q}$ with n sufficiently large. Hence,

$$E_{P^W}\left[\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\right|\right]\leq K'v^{1/2}\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}(-\ln\epsilon)d\epsilon=K'v^{1/2}(\tilde{\delta}_n-\tilde{\delta}_n\ln(\tilde{\delta}_n)). \tag{E.226}$$

By (E.224) and taking expectations with respect to P in (E.226), it follows that

$$P(Z_n(\tilde{\delta}_n)>\epsilon_n)\leq 2K'v^{1/2}(\tilde{\delta}_n-\tilde{\delta}_n\ln(\tilde{\delta}_n))/\epsilon_n=O(\delta_n^\gamma/\epsilon_n)+O(|\delta_n^\gamma\ln(\delta_n)|/\epsilon_n)=o(1), \tag{E.227}$$

where the last equality follows from the rate condition on ϵ_n. By (E.223) and (E.227), conclude that the first claim of the lemma holds.

For the second claim, define $Z_n^*(\tilde{\delta}_n) \equiv \sup_{f\in\mathcal{M}_{\tilde{\delta}_n}} |\sqrt{n}(\hat{P}_n^* - \hat{P}_n)f|$, where $\hat{P}_n^*$ is the empirical distribution of $\{X_i^b\}_{i=1}^n$. Then, by (E.221), one has

$$P_n^*\left(\sup_{\|\theta-\theta'\|\le\delta_n}\max_{j=1,\cdots,J}|\mathfrak{G}_{n,j}^b(\theta)-\mathfrak{G}_{n,j}^b(\theta')|>\epsilon_n\Big|\{X_i\}_{i=1}^\infty\right)\le P_n^*\big(Z_n^*(\tilde{\delta}_n)>\epsilon_n\big|\{X_i\}_{i=1}^\infty\big). \tag{E.228}$$

By Markov's inequality and Lemma 2.3.1 (symmetrization lemma) in van der Vaart and Wellner (2000), one has

$$P_n^*\big(Z_n^*(\tilde{\delta}_n)>\epsilon_n\big|\{X_i\}_{i=1}^\infty\big)\le\frac{2}{\epsilon_n}E_{P_n^*\times P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i^b)\Big|\Big|\{X_i\}_{i=1}^\infty\right] \tag{E.229}$$

$$=\frac{2}{\epsilon_n}E_{P_n^*}\left[E_{P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i^b)\Big|\Big|\{X_i^b\},\{X_i\}_{i=1}^\infty\right]\Big|\{X_i\}_{i=1}^\infty\right], \tag{E.230}$$

where $\{W_i\}_{i=1}^n$ are i.i.d. Rademacher random variables independent of $\{X_i\}_{i=1}^\infty$ and $\{M_{n,i}\}_{i=1}^n$. Argue as in (E.224)-(E.227). Then, it follows that

$$P_n^*(Z_n^*(\tilde{\delta}_n)>\epsilon_n|\{X_i\}_{i=1}^\infty)=O(\delta_n^\gamma/\epsilon_n)+O(-\delta_n^\gamma\ln(\delta_n)/\epsilon_n)=o(1),$$

for almost all sample paths. Hence, the second claim of the lemma follows. □

Lemma E.12: *Suppose Assumptions 4.1, 4.2, and 4.5 hold. Let* $\mathcal{S}_P\equiv\{f:\mathcal{X}\to\mathbb{R}:f(\cdot)=\sigma_{P,j}(\theta)^{-2}m_j^2(\cdot,\theta),\theta\in\Theta,j=1,\cdots,J\}$ *and let F be its envelope. (i) If $\mathcal{S}_P$ is Donsker and pre-Gaussian uniformly in $P\in\mathcal{P}$, then*

$$\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)|^*=O_{\mathcal{P}}(1/\sqrt{n}); \tag{E.231}$$

(ii) If $|\sigma_{P,j}(\theta)^{-1}m_j(x,\theta)-\sigma_{P,j}(\theta')^{-1}m_j(x,\theta')|\le\bar{M}(x)\|\theta-\theta'\|$ *with* $E_P[\bar{M}(X)^2]<M$ *for all* $\theta,\theta'\in\Theta$, $x\in\mathcal{X}$, $j=1,\cdots,J$, *and* $P\in\mathcal{P}$, *then, for any $\eta>0$, there exists a constant $C>0$ such that*

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}P\Big(\max_{j=1,\cdots,J}\sup_{\|\theta-\theta'\|<\delta}|\eta_{n,j}(\theta)-\eta_{n,j}(\theta')|>C\delta\Big)<\eta. \tag{E.232}$$

Proof. We show the claim by first showing that, for any $\delta>0$, there exist $M>0$ and $N\in\mathbb{N}$ such that

$$\inf_{P\in\mathcal{P}}P^\infty\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}-1\Big|\le M/\sqrt{n}\Big)\ge1-\delta,\ \forall n\ge N. \tag{E.233}$$

By Assumptions 4.1 (iv), 4.5 and Theorem 2.8.2 in van der Vaart and Wellner (2000), $\mathcal{M}_P$ is a Donsker class uniformly in $P\in\mathcal{P}$. By hypothesis, $\mathcal{S}_P$ is a Donsker class uniformly in $P\in\mathcal{P}$.

Therefore, by the continuous mapping theorem, for any $\epsilon>0$,

$$\Big|P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma_{P,j}^2(\theta)}-\frac{E_P[m_j(X,\theta)^2]}{\sigma_{P,j}^2(\theta)}\Big|\le C_1\Big)-\Pr(\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)|\le C_1)\Big|\le\epsilon \tag{E.234}$$

$$\Big|P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta)-E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big|\le C_2\Big)-\Pr(\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)|\le C_2)\Big|\le\epsilon. \tag{E.235}$$

for n sufficiently large uniformly in $P\in\mathcal{P}$, where $\mathbb{H}_{P,j}$ and $\mathbb{G}_{P,j}$ are tight Gaussian processes, and C_1 and C_2 are the continuity points of the distributions of $\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)|$ and $\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)|$ respectively. As in the proof of Lemma E.10 (i), bounding each term of the right hand side of (E.197) by $C_1/\sqrt{n}$ and $C_2/\sqrt{n}$ implies that

$\sup_{\theta\in\Theta}\left|\frac{\hat\sigma^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)}-1\right|\le C/\sqrt{n}$ for some constant $C>0$. Now choose $C_1>0$ and $C_2>0$ so that

$$\Pr(\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)|\le C_1)>1-\delta/3,\quad\text{and}\quad \Pr(\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)|\le C_2)>1-\delta/3, \tag{E.236}$$

and set $\epsilon>0$ sufficiently small so that $1-2\delta/3-2\epsilon\ge 1-\delta$. The existence of such continuity points $C_1,C_2>0$ is due to Theorem 11.1 in Davydov, Lifshitz, and Smorodina (1995) applied to $\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)|$ and $\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)|$ respectively. Then, for sufficiently large n,

$$\begin{aligned}1-\delta&\le P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma^2_{P,j}(\theta)}-\frac{E_P[m_j(X,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big|\le C_1,\\&\qquad\qquad\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{\bar m_{n,j}(\theta)-E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big|\le C_2\Big)\\&\le P\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat\sigma^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)}-1\Big|\le C/\sqrt{n}\Big),\end{aligned} \tag{E.237}$$

uniformly in $P\in\mathcal{P}$.

Next, note that, for $x>0$ and $0<\eta<1$, $|x^2-1|\le\eta$ implies $|x-1|\le 1-(1-\eta)^{1/2}\le\eta$, and hence by (E.237), for sufficiently large n,

$$1-\delta\le P\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat\sigma_{n,j}(\theta)}{\sigma_{P,j}(\theta)}-1\Big|\le C/\sqrt{n}\Big), \tag{E.238}$$

uniformly in $P\in\mathcal{P}$. Finally, note again that $|\hat\sigma_{n,j}(\theta)/\sigma_{P,j}(\theta)-1|\le\epsilon$ implies $\hat\sigma_{n,j}(\theta)>0$, and by the local Lipshitz continuity of $x\mapsto 1/x$ on a neighborhood around 1, there is a constant C' such that

$$P\Big(\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)|\le C'/\sqrt{n}\Big)\ge 1-\delta, \tag{E.239}$$

uniformly in $P\in\mathcal{P}$ for all n sufficiently large. This establishes the first claim of the lemma.

(ii) First, consider

$$\frac{\hat\sigma^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)}=n^{-1}\sum_{i=1}^n\left(\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2-\left(n^{-1}\sum_{i=1}^n\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2. \tag{E.240}$$

We claim that this function is Lipschitz with probability approaching 1. To see this, note that, for any $\theta,\theta'\in\Theta$,

$$\begin{aligned}&\left|n^{-1}\sum_{i=1}^n\left(\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2-n^{-1}\sum_{i=1}^n\left(\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right)^2\right|\\&\qquad=\left|n^{-1}\sum_{i=1}^n\left(\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}+\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right)\left(\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}-\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right)\right|\\&\qquad\le n^{-1}\sum_{i=1}^n 2\sup_{\theta\in\Theta}\left|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right|\bar M(X_i)\|\theta-\theta'\|.\end{aligned} \tag{E.241}$$

Define $B_n\equiv n^{-1}\sum_{i=1}^n 2\sup_{\theta\in\Theta}\left|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right|\bar M(X_i)$. By Markov and Cauchy-Schwarz inequalities,

$$P(B_n>C)\le\frac{E[B_n]}{C}\le\frac{2E_P\left[\sup_{\theta\in\Theta}\left|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right|^2\right]^{1/2}E_P\left[\bar M(X_i)^2\right]^{1/2}}{C}\le\frac{2M}{C}, \tag{E.242}$$

where the third inequality is due to Assumptions 4.1 (iv) and the assumption on $\bar{M}$. Hence, for any $\eta > 0$, one may find $C > 0$ such that $\sup_{P\in\mathcal{P}} P(B_n > C) < \eta$ for all n.

Similarly, for any $\theta, \theta' \in \Theta$,

$$\begin{aligned}
&\left|\left(n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2 - \left(n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right)^2\right| \\
&\quad = \left|n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)} + n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right|\left|n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)} - n^{-1}\sum_{i=1}^{n}\frac{m(X_i,\theta')}{\sigma_{P,j}(\theta')}\right| \\
&\quad \le n^{-1}\sum_{i=1}^{n} 2\sup_{\theta\in\Theta}\left|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right| n^{-1}\sum_{i=1}^{n}\bar{M}(X_i)\|\theta-\theta'\|. \qquad \text{(E.243)}
\end{aligned}$$

Define $\tilde{B}_n \equiv n^{-1}\sum_{i=1}^{n} 2\sup_{\theta\in\Theta}\left|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right| n^{-1}\sum_{i=1}^{n}\bar{M}(X_i)$. By Markov, Cauchy-Schwarz, and Jensen's inequalities,

$$\begin{aligned}
P(\tilde{B}_n > C) &\le \frac{E[\tilde{B}_n]}{C} \le \frac{2E_P\Big[\Big(n^{-1}\sum\sup_{\theta\in\Theta}\big|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\big|\Big)^2\Big]^{1/2} E_P\Big[\Big(n^{-1}\sum\bar{M}(X_i)\Big)^2\Big]^{1/2}}{C} \\
&\le \frac{2E_P\Big[\sup_{\theta\in\Theta}\big|\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\big|^2\Big]^{1/2} E_P[\bar{M}(X_i)^2]^{1/2}}{C} \le \frac{2M}{C}, \qquad \text{(E.244)}
\end{aligned}$$

where the last inequality is due to Assumptions 4.1 (iv) and the assumption on $\bar{M}$. Hence, for any $\eta > 0$, one may find $C > 0$ such that $\sup_{P\in\mathcal{P}} P(\tilde{B}_n > C) < \eta$ for all n.

Finally, let $g(y) \equiv y^{-1/2} - 1$ and note that $|g(y) - g(y')| \le \frac{1}{2}\sup_{\bar{y}\in(1-\epsilon,1+\epsilon)}|\bar{y}|^{-3/2}|y-y'|$ on $(1-\epsilon, 1+\epsilon)$. As shown in (E.238), $\hat{\sigma}^2_{n,j}(\theta)/\sigma^2_{P,j}(\theta)$ converges to 1 in probability, and g is locally Lipschitz on a neighborhood of 1. Combining this with (E.240)-(E.244) yields the desired result. □

Lemma E.13: *Suppose Assumption 4.1 holds. Suppose further that* $|\sigma_{P,j}(\theta)^{-1}m_j(x,\theta) - \sigma_{P,j}(\theta')^{-1}m_j(x,\theta')| \le \bar{M}(x)\|\theta-\theta'\|$ *with* $E_P[\bar{M}(X)^2] < M$ *for all* $\theta, \theta' \in \Theta$, $x \in \mathcal{X}$, $j = 1, \cdots, J$, *and* $P \in \mathcal{P}$.

Then,

$$\sup_{P\in\mathcal{P}}\|Q_P(\theta_1,\tilde{\theta}_1) - Q_P(\theta_2,\tilde{\theta}_2)\| \le M\|(\theta_1,\tilde{\theta}_1) - (\theta_2,\tilde{\theta}_2)\|, \qquad \text{(E.245)}$$

for some $M > 0$ *and for all* $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$.

Proof. Recall that

$$[Q_P(\theta_1,\tilde{\theta}_1)]_{j,k} = E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big]. \qquad \text{(E.246)}$$

For any $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$,

$$
\begin{aligned}
&\Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le \Big|E_P\Big[\Big(\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)} - \frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big)\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| + \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big(\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)} - \frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big)\Big]\Big| \\
&\le E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_k(X_i,\theta)}{\sigma_{P,k}(\theta)}\Big|\bar{M}(X_i)\Big]\|\theta_1-\theta_2\| + E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_j(X_i,\theta)}{\sigma_{P,j}(\theta)}\Big|\bar{M}(X_i)\Big]\|\tilde{\theta}_1-\tilde{\theta}_2\| \\
&\le M(\|\theta_1-\theta_2\| + \|\tilde{\theta}_1-\tilde{\theta}_2\|), \qquad \text{(E.247)}
\end{aligned}
$$

where the last inequality is due to the Cauchy-Schwarz inequality, Assumption 4.1 (iv), and the assumption on $\bar{M}$.

Similarly, for any $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$,

$$
\begin{aligned}
&\Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)} - \frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big]\Big|\Big|E_P\Big[\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| + \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]\Big|\Big|E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)} - \frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_k(X_i,\theta)}{\sigma_{P,k}(\theta)}\Big|\Big]E_P[\bar{M}(X_i)]\|\theta_1-\theta_2\| + E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_j(X_i,\theta)}{\sigma_{P,j}(\theta)}\Big|\Big]E_P[\bar{M}(X_i)]\|\tilde{\theta}_1-\tilde{\theta}_2\| \\
&\le M(\|\theta_1-\theta_2\| + \|\tilde{\theta}_1-\tilde{\theta}_2\|), \qquad \text{(E.248)}
\end{aligned}
$$

where the last inequality is due to the Cauchy-Schwarz inequality, Assumption 4.1 (iv), and the assumption on $\bar{M}$. The conclusion of the lemma then follows from (E.246)-(E.248). □

E.3 Almost Sure Representation Lemma and Related Results

In this appendix, we provide details on the almost sure representation used in Lemmas E.3, E.4, E.6, and E.9. We start with stating a uniform version of the bootstrap consistency in van der Vaart and Wellner (2000). For this, we define the original sample $X^\infty = (X_1, X_2, \cdots)$ and a n-dimensional multinomial vector M_n on a common probability space $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty) \times (\mathcal{Z}, \mathcal{C}, Q)$. We then view X^∞ as the coordinate projection on the first ∞ coordinates of the probability space above. Similarly, we view M_n as the coordinate projection on $\mathcal{Z}$. Here, M_n follows a multinomial distribution with parameter $(n; 1/n, \cdots, 1/n)$ and is independent of X^∞. We then let $E_M[\cdot|X^\infty = x^\infty]$ denote the conditional expectation of M_n given $X^\infty = x^\infty$. Throughout, we let $\ell^\infty(\Theta, \mathbb{R}^J)$ denote uniformly bounded $\mathbb{R}^J$-valued functions on Θ. We simply write $\ell^\infty(\Theta)$ when $J = 1$.

Using the multinomial weight, we rewrite the empirical bootstrap process as

$$
\mathbb{G}^b_{n,j}(\cdot) = g_j(X^\infty, M_n) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (M_{n,i} - 1) m_j(X_i, \cdot)/\hat{\sigma}_{n,j}(\cdot),\ j = 1, \cdots, J, \qquad \text{(E.249)}
$$

where $g_j : \mathcal{X}^\infty \times \mathcal{Z} \to \ell^\infty(\Theta)$ is a function that maps the sample path and the multinomial weight (X^∞, M_n) to the empirical bootstrap process $\mathbb{G}^b_{n,j}$. We then let $g : \mathcal{X}^\infty \times \mathcal{Z} \to \ell^\infty(\Theta, \mathbb{R}^J)$ be defined by $g = (g_1, \cdots, g_J)'$. For any function $f : \ell^\infty(\Theta, \mathbb{R}^J) \to \mathbb{R}$, the conditional expectation of $f(\mathbb{G}^b_n)$ given the sample path X^∞ is

$$
E_M[f(\mathbb{G}^b_n)|X^\infty = x^\infty] = \int f \circ g(x^\infty, m_n) dQ(m_n), \qquad \text{(E.250)}
$$

where, with a slight abuse of notation, we use Q for the induced law of M_n.

Let $\mathcal{F}$ be the function space $\{f(\cdot) = (m_1(\cdot,\theta)/\sigma_{P,1}(\theta), \cdots, m_J(\cdot,\theta)/\sigma_{P,J}(\theta)), \theta \in \Theta, P \in \mathcal{P}\}$. For each j, define a bootstrapped empirical process standardized by $\sigma_{P,j}$ as follows:

$$\mathfrak{G}^b_{n,j}(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(m_j(X_i^b,\theta) - \bar{m}_n(\theta)\right)/\sigma_{P,j}(\theta)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n (M_{n,i}-1)m_j(X_i,\theta)/\sigma_{P,j}(\theta). \tag{E.251}$$

The following result was shown in the proof of Lemma D.2.8 in Bugni, Canay, and Shi (2015), which is a uniform version of (a part of) Theorem 3.6.2 in van der Vaart and Wellner (2000). For the definition of a uniform version of Donskerness and pre-Gaussianity, we refer to van der Vaart and Wellner (2000) pages 168-169. Below, we let P^* denote the outer probability of P and let T^* denote the minimal measurable majorant of any (not necessarily measurable) random element T.

LEMMA E.14: *Let $\mathcal{F}$ be a class of measurable functions with finite envelope function. Suppose $\mathcal{F}$ is such that (i) $\mathcal{F}$ is Donsker and pre-Gaussian uniformly in $P \in \mathcal{P}$; and (ii) $\sup_{P\in\mathcal{P}} P^*\|f - Pf\|^2_{\mathcal{F}} < \infty$. Then,*

$$\sup_{h\in BL_1} |E_M[h(\mathfrak{G}^b_n)|X^\infty] - E[h(\mathbb{G}_P)]| \overset{as*}{\to} 0, \tag{E.252}$$

uniformly in $P \in \mathcal{P}$.

The result above gives uniform consistency of the standardized bootstrap process $\mathfrak{G}^b_n$. We now extend this to the studentized bootstrap process $\mathbb{G}^b_n$.

LEMMA E.15: *Suppose Assumptions 4.1, 4.2, and 4.5 hold. Then,*

$$\sup_{h\in BL_1} |E_M[h(\mathbb{G}^b_n)|X^\infty] - E[h(\mathbb{G}_P)]| \overset{as*}{\to} 0, \tag{E.253}$$

uniformly in $P \in \mathcal{P}$.

Proof. By Assumptions 4.1 (iv) and 4.5, Assumptions A.1-A.4 in Bugni, Canay, and Shi (2015) hold, which in turn implies that, by their Lemma D.1.2, $\mathcal{F}$ is Donsker and pre-Gaussian uniformly in $P \in \mathcal{P}$. Further, by Assumption 4.1 (iv) again, $\sup_{P\in\mathcal{P}} P^*\|f - Pf\|_{\mathcal{F}} < \infty$. Hence, by Lemma E.14,

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\sup_{h\in BL_1} |E_M[h(\mathfrak{G}^b_n)|X^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big) = 1. \tag{E.254}$$

For later use, we define the following set of sample paths, which has probability 1 uniformly in $P \in \mathcal{P}$.

$$A \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{h\in BL_1} |E_M[h(\mathfrak{G}^b_n)|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big\}. \tag{E.255}$$

Note that $\mathbb{G}^b_{n,j}$ and $\mathfrak{G}^b_{n,j}$ are related to each other by the following relationship:

$$\mathbb{G}^b_{n,j}(\theta) - \mathfrak{G}^b_{n,j}(\theta) = \mathfrak{G}^b_{n,j}(\theta)\left(\frac{\sigma_{P,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} - 1\right) = \mathfrak{G}^b_{n,j}(\theta)\eta_{n,j}(\theta),\ \theta \in \Theta. \tag{E.256}$$

By Assumptions 4.1, 4.2, and 4.5, Lemma E.10 applies. Hence,

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* \to 0\Big) = 1. \tag{E.257}$$

Define the following set of sample paths:

$$B \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* \to 0, \forall j = 1, \cdots, J\Big\}. \tag{E.258}$$

For any $x^\infty \in A \cap B$, it then follows that

$$\sup_{h\in BL_1} \big|E_M[h(\mathbb{G}_n^b)|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]\big|^* \to 0, \tag{E.259}$$

due to (E.254) and (E.256), h being Lipschitz, $\mathfrak{G}_{n,j}^b$ being bounded (given x^∞), and $\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* \to 0$ for all j. Finally, note that $\inf_{P\in\mathcal{P}} P^\infty(A \cap B) = 1$ due to (E.254), (E.257), and De Morgan's law. This establishes the conclusion of the lemma. □

The following lemma shows that, for almost all sample path x^∞, one can find an almost sure representation of the bootstrapped empirical process that is convergent.

LEMMA E.16: *Suppose Assumptions 4.1, 4.2, and 4.5 hold. Then, for each $x^\infty \in \mathcal{X}^\infty$, there exists a sequence $\{\tilde{G}_{n,x^\infty} \in \ell(\Theta, \mathbb{R}^J), n \geq 1\}$ and a random element $\tilde{G}_{P,x^\infty} \in \ell(\Theta, \mathbb{R}^J)$ defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbf{P}})$ such that*

$$\int h \circ g(x^\infty, m_n) dQ(m_n) = \int h(\tilde{G}_{n,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1 \tag{E.260}$$

$$\int h(\mathbb{G}_P(\omega)) dP(\omega) = \int h(\tilde{G}_{P,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1, \tag{E.261}$$

for all $x^\infty \in C$ for some set $C \subset \mathcal{X}^\infty$ such that $\inf_{P\in\mathcal{P}} P^\infty(C) = 1$ and

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\big\{x^\infty \in \mathcal{X}^\infty : \tilde{G}_{n,x^\infty} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{G}_{P,x^\infty}\big\}\Big) = 1. \tag{E.262}$$

Proof. Define the following set of sample paths:

$$C \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{h\in BL_1} |E_M[h(\mathbb{G}_{n,j}^b)|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big\}. \tag{E.263}$$

By Lemma E.15, $\inf_{P\in\mathcal{P}} P^\infty(C) = 1$.

For each fixed sample path $x^\infty \in C$, consider the bootstrap empirical process $g(x^\infty, M_n)$ in (E.249). This is a random element in $\ell^\infty(\Theta, \mathbb{R}^J)$ with a law governed by Q. For each $x^\infty \in C$, by Lemma E.15,

$$\sup_{h\in BL_1} \Big|\int h \circ g(x^\infty, m_n) dQ(m_n) - E[h(\mathbb{G}_P)]\Big|^* \to 0. \tag{E.264}$$

Hence, by Theorem 1.10.4 in van der Vaart and Wellner (2000), for each $x^\infty \in C$, one may find an almost sure representation $\tilde{G}_{n,x^\infty}$ of $g(x^\infty, M_n)$ on some probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbf{P}})$ such that

$$\int h \circ g(x^\infty, m_n) dQ(m_n) = \int h(\tilde{G}_{n,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1. \tag{E.265}$$

In particular, the proof of Theorem 1.10.4 in van der Vaart and Wellner (2000) (see also Addendum 1.10.5) allows us to take $\tilde{G}_{n,x^\infty}$ to be defined for each $\tilde{\omega} \in \tilde{\Omega}$ as

$$\tilde{G}_{n,x^\infty}(\tilde{\omega}) = g(x^\infty, M_n(\phi_n(\tilde{\omega}))), \tag{E.266}$$

for some perfect map $\phi_n : \tilde{\Omega} \to \mathcal{Z}$ (see the construction of ϕ_α in the middle of page 61 in VW). One may define $\tilde{G}_{n,x^\infty}$ arbitrarily for any $x^\infty \notin C$. The almost sure representation $\tilde{G}_{P,x^\infty}$ of $\mathbb{G}_{P,j}$ is defined similarly.

By Theorem 1.10.4 in van der Vaart and Wellner (2000), Eq. (E.259), and $\inf_{P\in\mathcal{P}} P(C)=1$, it follows that

$$\inf_{P\in\mathcal{P}} P^{\infty}\Big(\big\{x^{\infty}\in\mathcal{X}^{\infty}:\tilde{G}_{n,x^{\infty}} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{G}_{P,x^{\infty}}\big\}\Big)=1. \tag{E.267}$$

This establishes the claim of the lemma. □

Lemma E.17: *Suppose Assumptions 4.1, 4.2, and 4.5 hold. Let* $W_n\equiv(\mathbb{G}_n^b,Y_n)$ *be a sequence in* $\mathcal{W}\equiv\ell(\Theta,\mathbb{R}^J)\times\mathbb{R}^{d_Y}$ *such that* $Y_n=\tilde{g}(X^{\infty},M_n)$ *for some map* $\tilde{g}:\mathcal{X}^{\infty}\times\mathcal{Z}\to\mathbb{R}^{d_Y}$ *and*

$$\inf_{P\in\mathcal{P}} P^{\infty}\big(\sup_{h\in BL_1}|E_M[h(W_n)|X^{\infty}=x^{\infty}]-E[h(W)]|^{*}\to 0\big)=1, \tag{E.268}$$

where $W=(\mathbb{G},Y)$ *is a Borel measurable random element in* $\mathcal{W}$.

Then, for each $x^{\infty}\in\mathcal{X}^{\infty}$, *there exists a sequence* $\{W^{*}_{n,x^{\infty}}\in\mathcal{W},n\ge 1\}$ *and a random element* $W^{*}_{x^{\infty}}\in\mathcal{W}$ *defined on some probability space* $(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{\mathbf{P}})$ *such that*

$$E_M[h(W_n)|X^{\infty}=x^{\infty}]=\int h(W^{*}_{n,x^{\infty}}(\tilde{\omega}))d\tilde{\mathbf{P}}^{*}(\tilde{\omega}),\ \forall h\in BL_1 \tag{E.269}$$

$$E[h(W)]=\int h(W^{*}_{x^{\infty}}(\tilde{\omega}))d\tilde{\mathbf{P}}^{*}(\tilde{\omega}),\ \forall h\in BL_1, \tag{E.270}$$

for all $x^{\infty}\in C$ *for some set* $C\subset\mathcal{X}^{\infty}$ *such that* $\inf_{P\in\mathcal{P}}P^{\infty}(C)=1$, *and*

$$\inf_{P\in\mathcal{P}} P^{\infty}\Big(\big\{x^{\infty}\in\mathcal{X}^{\infty}:W^{*}_{n,x^{\infty}} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{W}^{*}_{x^{\infty}}\big\}\Big)=1. \tag{E.271}$$

Proof. Let $C\equiv\{x^{\infty}:\sup_{h\in BL_1}|E_M[h(W_n)|X^{\infty}=x^{\infty}]-E[h(W)]|^{*}\to 0\}$. The rest of the proof is the same as the one for Lemma E.16 and is therefore omitted. □

Remark E.1: When called by the Lemmas in Appendix E, Lemma E.17 is applied, for example, with $Y_n=(vec(\hat{D}_n(\theta'_n)),\hat{\xi}_n(\theta'_n))$ and $Y=(vec(D),\pi_1)$.

Appendix F Further Comparison of Calibrated Projection and BCS-Profiling

We next show that finite sample power can be higher with calibrated projection than with BCS-profiling, and that, due to the slow rate at which κ_n diverges, this effect can be large in samples of considerable size. Thus, the approaches are not nested in terms of power in empirically relevant examples. We then provide an example where all of calibrated projection, BCS-profiling and the method of Pakes, Porter, Ho, and Ishii (2011) fail in a specific instance where Assumption 4.3 is not satisfied.

F.1 Finite Sample Comparison in a Specific Example

We next analyze a stylized example of one-sided testing when the support set in direction p is a singleton identified as the intersection of d moment inequalities with regular geometry. In this example, calibrated projection has more power (less false coverage) than BCS-profiling, and the numerical difference can be large. The example resembles empirically important cases, namely polyhedral identified sets with large interior, e.g. linear regression with interval outcome data; recall that by Theorem 4.3, the two-sided testing problem reduces to two one-sided ones in these cases. At the same time, we emphasize that other examples will go the other way, especially as the present example utilizes the simplifications from Theorem 4.3 and therefore has no ρ-box.

Let θ be partially identified by moment conditions

$$E_P(z^{j\prime}\theta - X_j) \leq 0, j = 1, \ldots, d.$$

Note that to simplify the analysis, we assume exactly d conditions. Assume that $\{z^1, \ldots, z^d\}$ are linearly independent and also that p is in their positive span, so that Θ_I is bounded in direction p but not $-p$. The confidence intervals will be accordingly one-sided. Since gradients are known, all simplifications from Theorem 4.3 apply. We borrow from algebra in the proof of Theorem 4.4 to observe that, with the simplifications in place, CI_n and CI_n^{prof} invert tests that use the same test statistic but different bootstrap approximations to its distribution as follows:

$$\begin{aligned} T_n^{DR} &= \max_j \left\{\mathbb{G}_{n,j}^b\right\} \\ T_n^{PR}(s_n) &= \min_{p'\lambda=0} \max_j \left\{ \mathbb{G}_{n,j}^b + \underbrace{\frac{\sqrt{n}}{\kappa_n} \frac{z^{j\prime}\hat{\theta}_{p,s_n}^* - \bar{X}_j}{\hat{\sigma}_{n,j}}}_{>0} + \frac{z^{j\prime}\lambda}{\hat{\sigma}_{n,j}} \right\} \\ T_n^b &= \min_{p'\lambda=0} \max_j \left\{ \mathbb{G}_{n,j}^b + \underbrace{\frac{\sqrt{n}}{\kappa_n} \frac{z^{j\prime}\hat{\theta}_p^* - \bar{X}_j}{\hat{\sigma}_{n,j}}}_{=0} + \frac{z^{j\prime}\lambda}{\hat{\sigma}_{n,j}} \right\} \leq \min\{T_n^{DR}, T_n^{PR}(s_n)\}, \end{aligned}$$

where (as in Theorem 4.3) s_n is the value of $p'\theta$ being tested and $\hat{\theta}_{p,s_n}$ minimizes the sample criterion subject to $p'\theta = s_n$. The last inequality is strict unless the problem defining T_n^b is solved by $\lambda = 0$. The assessments of intercept terms in $T_n^{PR}(s_n)$ and T_n^b use that by construction of the example, all sample constraints bind at $\hat{\theta}_p^*$ and are violated at $\hat{\theta}_{p,s_n}^*$ (else, the test statistic would be 0 and the critical value not computed). Equality thus requires knife-edge realizations of $\mathbb{G}_{n,j}^b$, so its probability vanishes as $\mathbb{G}_{n,j}^b$ approaches multivariate normality and is in fact 0 for typical empirical samples. We conclude that the calibrated projection CI_n is deterministically a weak (and essentially always a strict) subset of the BCS-profiling CI_n^{prof} in this example.

We next provide a numerical comparison in a further stripped-down version of the example. Thus, consider one-sided testing with moment conditions

$$\begin{aligned} -\theta_1 + \theta_2 - E_P(X_1) &\leq 0 \\ \theta_1 + \theta_2 - E_P(X_2) &\leq 0 \end{aligned}$$

where the data are $(X_1, X_2) \sim N((E_P(X_1), E_P(X_2)), I_2)$ and $E_P(X_1) = E_P(X_2) = 0$. All of these facts other than $E_P(X_1) = E_P(X_2) = 0$, but including the gradients and variance matrix, are known. This enables closed form arguments. Also, for a researcher knowing this, the natural bootstrap implementation is a parametric bootstrap:

$$\begin{aligned} &\left(X_1^b, X_2^b\right) \sim N((\bar{X}_1, \bar{X}_2), I_2) \\ \Longrightarrow\quad &\sqrt{n}(\bar{X}_1^b - \bar{X}_1, \bar{X}_2^b - \bar{X}_1) = (Z_1, Z_2) \sim N(0, I_2) \end{aligned}$$

which we will use, i.e. (Z_1, Z_2) will take the role of $(\mathbb{G}_{n,1}^b, \mathbb{G}_{n,2}^b)$. Numerical computations refer to $\alpha = 5\%$.

Let $p = (0, 1)$. We construct one-sided confidence intervals for $s(p, \Theta_I(P))$. All intervals contain $(-\infty, s(p, \hat{\Theta}_I)]$, and simple algebra shows $s(p,\hat{\Theta}_I)=\frac{\bar{X}_1+\bar{X}_2}{2}$. Also noting that in this example $s(p,\Theta_I(P))=0$ and, for $s_n>s(p,\hat{\Theta}_I)$,

$$\begin{aligned}
H(p,\hat{\Theta}_I) &= \left\{\left(\frac{-\bar{X}_1+\bar{X}_2}{2},\frac{\bar{X}_1+\bar{X}_2}{2}\right)\right\}\\
\hat{\Theta}_I(s_n) &\equiv \left\{\theta\in\Theta: p'\theta=s_n,\, Q_n(\theta)\le \inf_{\theta\in\Theta:p'\theta=s_n} Q_n(\theta)\right\}=\left\{\left(\frac{-\bar{X}_1+\bar{X}_2}{2},s_n\right)\right\}\\
T_n(s_n) &= \sqrt{n}\max\left\{s_n-\frac{\bar{X}_1+\bar{X}_2}{2},0\right\},
\end{aligned}$$

where $Q_n(\theta)=\max_{j=1,\dots,J_1}(\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta))_+$, we compute

$$\begin{aligned}
T_n^{DR} &= \min_{\theta\in\hat{\Theta}_I(s_n)}\max\left\{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right),\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right),0\right\}\\
&= \max\left\{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right),\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right),0\right\}\sim\max\{Z_1,Z_2,0\}\\
T_n^{PR}(s_n) &= \min_{\theta_1\in\mathbb{R}}\max\left\{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right)+\kappa_n^{-1}\sqrt{n}\left(-\theta_1+s_n-\bar{X}_1\right),\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)+\kappa_n^{-1}\sqrt{n}\left(\theta_1+s_n-\bar{X}_2\right),0\right\}.
\end{aligned}$$

Unless its value is 0, the minimization problem defining $T_n^{PR}(s_n)$ is solved by setting two terms equal:

$$\theta_1=\frac{\sqrt{n}\left(\bar{X}_1^b-\hat{\mu}_1\right)-\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)+\kappa_n^{-1}\sqrt{n}\left(\bar{X}_2-\bar{X}_1\right)}{2\kappa_n^{-1}\sqrt{n}},$$

leading to

$$\begin{aligned}
T_n^{PR}(s_n) &= \max\left\{\frac{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right)+\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)}{2}+\kappa_n^{-1}\sqrt{n}\left(s_n-\frac{\bar{X}_1+\bar{X}_2}{2}\right),0\right\}\\
&= \max\left\{\frac{Z_1+Z_2}{2}+\kappa_n^{-1}\sqrt{n}\left(s_n-\frac{\bar{X}_1+\bar{X}_2}{2}\right),0\right\}=\max\left\{\frac{Z_1+Z_2}{2}+\kappa_n^{-1}T_n(s_n),0\right\}.
\end{aligned}$$

Finally, very similar reasoning to the above gives

$$\begin{aligned}
T_n^b &= \min_{\lambda\in\mathbb{R}}\max\left\{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right)+\kappa_n^{-1}\sqrt{n}\min\left(\frac{\bar{X}_1-\bar{X}_2}{2}+\frac{\bar{X}_1+\bar{X}_2}{2}-\bar{X}_1,0\right)-\lambda,\right.\\
&\qquad\left.\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)+\kappa_n^{-1}\sqrt{n}\min\left(\frac{-\bar{X}_1+\bar{X}_2}{2}+\frac{\bar{X}_1+\bar{X}_2}{2}-\bar{X}_2,0\right)+\lambda,0\right\}\\
&= \min_{\lambda\in\mathbb{R}}\max\left\{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right)-\lambda,\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)+\lambda,0\right\}\\
&= \max\left\{\frac{\sqrt{n}\left(\bar{X}_1^b-\bar{X}_1\right)+\sqrt{n}\left(\bar{X}_2^b-\bar{X}_2\right)}{2},0\right\}\\
&= \max\left\{\frac{Z_1+Z_2}{2},0\right\}.
\end{aligned}$$

Thus calibrated projection yields a critical value of $\hat{c}_n=\Phi^{-1}(1-\alpha)/\sqrt{2}\approx 1.16$, whereas simple projection uses $\hat{c}_n^{proj}=\Phi^{-1}(\sqrt{1-\alpha})\approx 1.95$; both are independent of s_n as well as n. BCS-profiling uses a critical value $\hat{c}_n^{prof}(s_n)$ that increases in the test statistic (hence, conditional on the data, in s_n) because the statistic itself enters T_n^{PR}. To facilitate a comparison, one can compute the fixed point at which $T_n(s_n)=\hat{c}_n^{prof}(s_n)$. BCS-profiling is equivalent to comparing $T_n(s_n)$ to that fixed point at all s_n, and we will therefore equate it with use of this critical value, labeled $\tilde{c}_n^{prof}$ below. This critical value converges to $\hat{c}_n$ at a rate of κ_n^{-1}, illustrating asymptotic equivalence of inference methods off the null in this case. However, for the popular choice of $\kappa_n=\sqrt{\log n}$, convergence is so slow that it should not be taken to describe behavior at realistic sample sizes. Table F.1 displays the numerical

Table F.1: Finite sample noncoverage rates in a specific example.

Type of cv	**n**	**Value**	**Power at $\gamma n^{-1/2}$, $\gamma = \ldots$** 0	1	2	3	4
$\hat{c}_n^{proj}$	any	1.95	.003	.089	.523	.930	.998
$\tilde{c}_n^{prof}$	10^3	1.63	.011	.188	.701	.974	1.000
$\tilde{c}_n^{prof}$	10^5	1.52	.016	.231	.751	.982	1.000
$\tilde{c}_n^{prof}$	10^7	1.47	.019	.254	.774	.985	1.000
$\tilde{c}_n^{prof}$	10^9	1.43	.022	.271	.790	.987	1.000
$\tilde{c}_n^{prof}$	10^{11}	1.40	.024	.284	.800	.988	1.000
$\tilde{c}_n^{prof}$	10^{13}	1.38	.025	.292	.807	.989	1.000
$\tilde{c}_n^{prof}$	10^{15}	1.37	.026	.299	.813	.989	1.000
$\tilde{c}_n^{prof}$	10^{17}	1.36	.027	.307	.819	.990	1.000
$\tilde{c}_n^{prof}$	10^{19}	1.35	.028	.313	.823	.990	1.000
$\tilde{c}_n^{prof}$	10^{50}	1.28	.036	.348	.847	.993	1.000
$\tilde{c}_n^{prof}$	10^{100}	1.24	.039	.366	.858	.994	1.000
$\hat{c}_n$	any	1.16	.050	.409	.882	.995	1.000

value of $\tilde{c}_n^{prof}$ and the implied noncoverage probability (or power) at $\gamma/\sqrt{n}$ for $\gamma \in \{0,1,2,3,4\}$; note that $\gamma = 0$ corresponds to the true support function. By construction, $\tilde{c}_n^{prof}$ interpolates between $\hat{c}_n^{proj}$ and $\hat{c}_n$ in this example, but convergence to $\hat{c}_n$ requires extreme sample sizes. For example, on the boundary edge of the true projection $CI_{.95}^{prof}$ has finite sample coverage of .975, which is effectively halfway between projection and calibrated projection, for $n = 10^{13}$.

F.2 Example of Methods Failure When Assumption 4.3 Fails

Consider one-sided testing with two inequality constraints in $\mathbb{R}^2$. The constraints are

$$\begin{aligned} \theta_1 + \theta_2 &\leq E_P(X_1) \\ \theta_1 - \theta_2 &\leq E_P(X_2). \end{aligned}$$

The projection of $\Theta_I(P)$ in direction $p = (1,0)$ is $(-\infty, (E_P(X_1) + E_P(X_2))/2]$, the support set is $H(p,\Theta_I) = \{((E_P(X_1)+E_P(X_2))/2, (E_P(X_1)-E_P(X_2))/2)\}$, and the support function takes value $\theta_1^* = (E_P(X_1)+E_P(X_2))/2$.

The random variables $(X_1, X_2)'$ have a mixture distribution as follows:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \begin{cases} N\left(0, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right) & \text{with probability } 1 - 1/n, \\ \delta_{(1,1)} \text{ (degenerate)} & \text{otherwise,} \end{cases}$$

hence $E_P(X_1) = E_P(X_2) = \theta_1^* = 1/n$. Note in particular the implication that

$$\frac{X_1 + X_2}{2} = \begin{cases} 0 & \text{with probability } 1 - 1/n, \\ 1 & \text{otherwise.} \end{cases}$$

The natural estimator of θ_1^* is $\hat{\theta}_1^* = (\bar{X}_1 + \bar{X}_2)/2$. It is distributed as Z/n, where Z is Binomial with parameters $(1/n, n)$. For large n, the distribution of Z is well approximated as Poisson with parameter 1. In particular, with probability approximately $e^{-1} \approx 37\%$, every sample realization of $(X_1 + X_2)/2$ equals zero. In this case, the following happens: (i) The projection of the sample analog of the identified set is $(-\infty, 0]$, so that a strictly positive critical value or level would be needed to cover the true projection. (ii) Because the empirical distribution of $(X_1 + X_2)/2$ is degenerate at zero, the distribution of $(\bar{X}_1^b + \bar{X}_2^b)/2$ is as well. Hence, all of Pakes, Porter, Ho, and Ishii (2011), Bugni, Canay, and Shi (2017), and calibrated projection (each with either parametric or nonparametric bootstrap) compute critical values or relaxation levels of 0.

This bounds from above the true coverage of all of these methods at $e^{-1} \approx 63\%$. Note that $(m < n)$-subsampling will encounter the same problem. Next we provide some discussion of the example.

Violation of Assumptions. The example violates our Assumption 4.3 because $Cov(X_1, X_2) \to 1$. It also violates Assumption 2 in Bugni, Canay, and Shi (2017): Their Assumption A2-(b) should apply, but the profiled test statistic on the true null concentrates at $1/n$. The example satisfies the assumptions explicitly stated in Pakes, Porter, Ho, and Ishii (2011), illustrating an oversight in their Theorem 2. (We here refer to the inference part of their 2011 working paper. We identified corresponding oversights in the proof of their Proposition 6.)

The example satisfies the assumptions of Andrews and Soares (2010) and Andrews and Guggenberger (2009), and both methods work here. The reason is that both focus on the distribution of the criterion function at a fixed θ and are not affected by the irregularity of $\hat{\theta}_1^*$.

Relation to Mammen (1992). In this example, all of Bugni, Canay, and Shi (2017), Pakes, Porter, Ho, and Ishii (2011), and our calibrated projection method reduce to one-sided nonparametric percentile bootstrap confidence intervals for $(E_P(X_1) + E_P(X_2))/2$ estimated by $(\bar{X}_1 + \bar{X}_2)/2$. By Mammen (1992, Theorem 1), asymptotic normality of an appropriately standardized estimator, i.e.

$$\exists \{a_n\} : a_n \left((\bar{X}_1 + \bar{X}_2) - (E_P(X_1) + E_P(X_2)) \right) \xrightarrow{d} N(0, 1),$$

is *necessary and* sufficient for this interval to be valid. This fails (the true limit is recentered Poisson at rate $a_n = n$), so that validity of any of the aforementioned methods would contradict the Theorem.

Appendix G Comparison with Projection of AS

In this Appendix we establish that for each $n \in \mathbb{N}$, CI_n is a subset of a confidence interval obtained by projecting an AS confidence set and denoted CI_n^{proj}.[8] Moreover, we derive simple conditions under which our confidence interval is a proper subset of the projection of AS's confidence set. Below we let $\hat{c}_n^{proj}$ denote the critical value obtained applying AS with criterion function $Q_n(\theta) = \max \{\max_{j=1,\ldots,J_1} (\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta))_+, \max_{j=J_1+1,\cdots,J_1+J_2} |\sqrt{n}\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)|\}$ and with the same choice as for $\hat{c}_n$ of GMS function φ and tuning parameter κ_n.

[8] Of course, AS designed their confidence set to uniformly cover each vector in Θ_I with prespecified asymptotic probability, a different inferential problem than the one considered here.

THEOREM G.1: *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, and 4.5 hold. Let* $0 < \alpha < 1/2$. *Then for each* $n \in \mathbb{N}$

$$CI_n \subseteq [-s(-p, \mathcal{C}_n(\hat{c}_n^{proj})), s(p, \mathcal{C}_n(\hat{c}_n^{proj}))], \tag{G.1}$$

where for given function c, $\mathcal{C}_n(c)$ *is defined in* (1.1)

Proof. For given θ, the event

$$\max_{j=1,\dots,J} \left\{ \mathbb{G}_{n,j}^b(\theta) + \varphi_j(\hat{\xi}_{n,j}(\theta)) \right\} \le c \tag{G.2}$$

implies the event

$$\max_{\lambda \in \Lambda_n^b(\theta,\rho,c)} p'\lambda \ge 0 \ge \min_{\lambda \in \Lambda_n^b(\theta,\rho,c)} p'\lambda, \tag{G.3}$$

with Λ_n^b defined in (3.1). This is so because if $\max_{j=1,\dots,J} \left\{ \mathbb{G}_{n,j}^b(\theta) + \varphi_j(\hat{\xi}_{n,j}(\theta)) \right\} \le c$, $\lambda = 0$ is feasible in both optimization problems in (G.3), hence the event in (G.3) is implied. In turn this yields that for each $n \in \mathbb{N}$ and $\theta \in \Theta$,

$$c_n^{proj}(\theta) \ge \hat{c}_n(\theta), \tag{G.4}$$

and therefore the result follows. □

The result in Theorem G.1 is due to the following fact. Recall that AS's confidence region calibrates its critical value so that, at each θ, the following event occurs with probability at least $1 - \alpha$:

$$\max_{j=1,\dots,J} \left\{ \mathbb{G}_{n,j}^b(\theta) + \varphi_j(\hat{\xi}_{n,j}(\theta)) \right\} \le c. \tag{G.5}$$

A natural question is, then, whether there are conditions under which CI_n is strictly shorter than the projection of AS's confidence region. Heuristically, this is the case with probability approaching 1 when $\hat{c}_n(\theta)$ is strictly less than $\hat{c}_n^{proj}(\theta)$ at each θ that is relevant for projection. For this, restrict $\varphi(\cdot)$ to satisfy $\varphi_j(x) \le 0$ for all x, fix θ and consider the pointwise limit of (G.5):

$$\mathbb{G}_{P,j}(\theta) + \zeta_{P,j}(\theta) \le c, \ \ j = 1, \cdots, J, \tag{G.6}$$

where $\{\mathbb{G}_{P,j}(\theta), j = 1, \cdots, J\}$ follows a multivariate normal distribution, and $\zeta_{P,j}(\theta) \equiv (-\infty)\mathbf{1}(\sqrt{n}\gamma_{1,P,j}(\theta) < 0)$ is the pointwise limit of $\varphi_j(\hat{\xi}_{n,j}(\theta))$ (with the convention that $(-\infty)0 = 0$). Under mild regularity conditions, $\hat{c}_n^{proj}(\theta)$ then converges in probability to a critical value $c = c^{proj}(\theta)$ such that (G.6) holds with probability $1-\alpha$. Similarly, the limiting event that corresponds to our problem (3.4) is

$$\Lambda(\theta, \rho, c) \cap \{p'\lambda = 0\} \ne \emptyset, \tag{G.7}$$

where the limiting feasibility set $\Lambda(\theta, \rho, c)$ is given by

$$\Lambda(\theta, \rho, c) = \{\lambda \in \rho B_{n,\rho}^d : \mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)\lambda + \zeta_{P,j}(\theta) \le c, j = 1, \cdots, J\}. \tag{G.8}$$

Note that if the gradient $D_{P,j}(\theta)$ is a scalar multiple of p, i.e. $D_{P,j}(\theta)/\|D_{P,j}(\theta)\| \in \{p, -p\}$, for all j such that $\zeta_{P,j}(\theta) = 0$, the two problems are equivalent because (G.6) implies (G.7) (by arguing that $\lambda = 0$ is in $\Lambda(\theta, \rho, c)$), and for the converse implication, whenever (G.7) holds, there is λ such that $\mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)\lambda + \zeta_{P,j}(\theta) \le c$ and $p'\lambda = 0$. Since $D_{P,j}(\theta)\lambda = 0$ for all j such that $\zeta_{P,j}(\theta) = 0$, one has $\mathbb{G}_{P,j}(\theta) + \zeta_{P,j}(\theta) \le c$ for all j.[9] In this special

[9]The gradients of the non-binding moment inequalities do not matter here because $\mathbb{G}_{P,j}(\theta) + \zeta_{P,j}(\theta) \le c$ holds due to $\zeta_{P,j}(\theta) = -\infty$ for such constraints.

case, the limits of the two critical values coincide asymptotically, but any other case is characterized by projection conservatism. Lemma G.1 below formalizes this insight. Specifically, for fixed θ, the limit of $\hat{c}_n(\theta)$ is strictly less than the limit of $\hat{c}_n^{proj}(\theta)$ if and only if there is a constraint that binds or is violated at θ and has a gradient that is not a scalar multiple of p.[10]

The parameter values that are relevant for the lengths of the confidence intervals are the ones whose projections are in a neighborhood of the projection of the identified set. Therefore, a leading case in which our confidence interval is strictly shorter than the projection of AS asymptotically is that in which at any θ (in that neighborhood of the projection of the identified set) at least one local-to-binding or violated constraint has a gradient that is not parallel to p. We illustrate this case with an example based on Manski and Tamer (2002).

Example G.1 (Linear regression with an interval valued outcome): Consider a linear regression model:

$$E[Y|Z] = Z'\theta, \tag{G.9}$$

where Y is an unobserved outcome variable, which takes values in the interval $[Y_L, Y_U]$ with probability one, and Y_L, Y_U are observed. The vector Z collects regressors taking values in a finite set $S_Z \equiv \{z_1, \cdots, z_K\}, K \in \mathbb{N}$. We then obtain the following conditional moment inequalities:

$$E_P[Y_L|Z = z_j] \leq z_j'\theta \leq E_P[Y_U|Z = z_j], \; j = 1, \cdots, K, \tag{G.10}$$

which can be converted into unconditional moment inequalities with $J_1 = 2K$ and

$$m_j(X, \theta) = \begin{cases} Y_L 1\{Z = z_j\}/g(z_j) - z_j'\theta, & j = 1, \cdots, K \\ z_{j-K}'\theta - Y_U 1\{Z = z_{j-K}\}/g(z_{j-K}) & j = K+1, \cdots, 2K, \end{cases} \tag{G.11}$$

where g denotes the marginal distribution of Z, which is assumed known for simplicity. Consider making inference for the value of the regression function evaluated at a counterfactual value $\tilde{z} \notin S_Z$. Then, the projection of interest is $\tilde{z}'\theta$. Note that the identified set is a polyhedron whose gradients are given by $D_{P,j}(\theta) = -z_j/\sigma_{P,j}, j = 1, \cdots, K$ and $D_{P,j}(\theta) = z_{j-K}/\sigma_{j-K}, j = K+1, \cdots, 2K$. This and $\tilde{z} \notin S_Z$ imply that for any θ not in the interior of the identified set, there exists a binding or violated constraint whose gradient is not a scalar multiple of p. Hence, for all such θ, our critical value is strictly smaller than $c_n^{proj}(\theta)$ asymptotically. In this case, our confidence interval becomes strictly shorter than that of AS asymptotically. We also note that the same argument applies even if the marginal distribution of Z is unknown. In such a setting, one needs to work with a sample constraint of the form $n^{-1}\sum_{i=1}^n Y_{L,i} 1\{Z_i = z_j\}/n^{-1}\sum_{i=1}^n 1\{Z_i = z_j\} - z_j\theta$ (and similarly for the upper bound). This change only alters the (co)variance of the Gaussian process in our limiting approximation but does not affect any other term.

We now provide a numerical illustration for a further simplified example. Assume that $p = (d^{-1/2}, ..., d^{-1/2}) \in \mathbb{R}^d$ and that there are d binding moment inequalities whose gradients are known and correspond to rows of the identity matrix. Assume furthermore that $\mathbb{G}$ is known to be exactly d-dimensional multivariate standard Normal. (Thus, Θ_I is the negative quadrant. Its unboundedness from below is strictly for simplicity.) Also, by Theorem 4.3, one can set $\rho = +\infty$ in this example.

Under these simplifying assumptions (which can, of course, be thought of as asymptotic approximations), it is

[10] The condition that all binding moment inequalities have gradient collinear with p is not as exotic as one might think. An important case where it obtains is the "smooth maximum," i.e. the support set is a point of differentiability of the boundary of Θ_I.

Table G.1: Conservatism from projection in a one-sided testing problem as a function of d

d	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**100**	∞
$\hat{c}_n$	1.64	1.16	0.95	0.82	0.74	0.67	0.62	0.58	0.55	0.52	0.16	0
$\hat{c}_n^{proj}$	1.64	1.95	2.12	2.23	2.32	2.39	2.44	2.49	2.53	2.57	3.28	∞
$1-\alpha^*$	.95	.77	.57	.40	.27	.18	.11	.07	.04	.03	10^{-25}	0

easy to calculate in closed form that

$$\begin{aligned} \hat{c}_n &= d^{-1/2}\Phi^{-1}(1-\alpha), \\ \hat{c}_n^{proj} &= \Phi^{-1}\left((1-\alpha)^{1/d}\right). \end{aligned}$$

Furthermore, for any $\alpha < 1/2$, one can compute α^* s.t. applying $\hat{c}_n$ with target coverage $(1-\alpha)$ yields the same confidence interval as using $\hat{c}_n^{proj}$ with target coverage $(1-\alpha^*)$.[11] Some numerical values are provided in Table G.1 (with $\alpha = 0.05$).

To cover $p'\theta$ in $\mathbb{R}^{10}$ with probability 95%, it suffices to project an AS-confidence region of size 3%. The example is designed to make a point; our Monte Carlo analyses in Section 5 showcase less extreme cases. However, the core defining feature of the example – namely, the identified set has a thick interior, and the support set is the intersection of d moment inequalities – frequently occurs in practice, and all such examples will qualitatively resemble this one as d grows large.

G.1 Necessary and Sufficient Condition for $\hat{c}_n(\theta) < \hat{c}_n^{proj}(\theta)$

The following lemma establishes the effect of ρ on $\hat{c}_n(\theta)$. In doing so it establishes a necessary and sufficient condition for $\hat{c}_n(\theta) < \hat{c}_n^{proj}(\theta)$, because the latter can be seen as the former calibrated with ρ set equal to zero. The lemma requires $\varphi_j(x) \leq 0$ for all x.[12]

Lemma G.1: *Fix $\theta \in \Theta$, $P \in \mathcal{P}$ and a value $\rho \in \mathbb{R}_+$. Suppose Assumptions 4.1, 4.2, 4.3, 4.4 and 4.5 hold and also that $\varphi_j(x) \leq 0$ for all x and j. Let $0 < \delta < \rho$. For $n \geq N$, let $\hat{c}_n(\theta)$ be calibrated using ρ in place of ρ, which necessarily yields a larger value for $\hat{c}_n(\theta)$. With a modification of notation, explicitly highlight $\hat{c}_n(\theta)$'s dependence on ρ through the notation $\hat{c}_n(\theta, \rho)$. Then*

$$|\hat{c}_n(\theta,\rho) - \hat{c}_n(\theta,\rho-\delta)| \xrightarrow{p} 0 \tag{G.12}$$

if and only if $D_{P,j}(\theta)/\|D_{P,j}(\theta)\| \in \{p, -p\}$ for all $j \in \mathcal{J}^*(\theta) \equiv \{j : E_P[m_j(X_i,\theta)] \geq 0\}$.

Remark G.1: For θ such that $\mathcal{J}^*(\theta) = \emptyset$, we have $\hat{c}_n(\theta,\rho) \xrightarrow{p} 0$ but also $\hat{c}_n^{proj}(\theta) \xrightarrow{p} 0$. This is consistent with Lemma G.1 because the condition on gradients vacuously holds in this case.

[11] Equivalently, $(1-\alpha^*)$ is the probability that $\mathcal{C}_n(\hat{c}_n^{proj})$ contains $\{0\}$, the true support set in direction p which furthermore, in this example, minimizes coverage within $\Theta_I(P)$. The closed-form expression is $1-\alpha^* = \Phi(d^{-1/2}\Phi^{-1}(1-\alpha))^d$. AS prove validity of their method only for $\alpha < 1/2$, but this is not important for the point made here.

[12] To keep the treatment general, we have not imposed this restriction throughout the paper. However, we only recommend functions φ_j with this feature anyway: for any φ_j that can take strictly positive values, substituting $\min\{\varphi_j(x), 0\}$ attains the same asymptotic size but generates CIs that are weakly shorter for all and strictly shorter for some sample realizations.

Proof. Recall that θ and P are fixed, i.e. we assume a pointwise perspective. Then

$$\hat{c}_n(\theta,\rho) \xrightarrow{p} \inf\{c \geq 0 : P(\{\lambda \in \rho B^d_{n,\rho} : \mathbb{G}_{P,j}(\theta) + D_{P,j}(\theta)\lambda \leq c, j \in \mathcal{J}^*(\theta)\} \cap \{p'\lambda = 0\} \neq \emptyset) \geq 1-\alpha\}. \tag{G.13}$$

Here, we used convergence of $\mathbb{G}^b_j(\theta)$ to $\mathbb{G}_{P,j}(\theta)$ and of $\hat{D}_j(\theta)$ to $D_{P,j}(\theta)$, boundedness of gradients, and the fact that

$$\varphi_j(\kappa_n^{-1}\sqrt{n}\bar{m}_j(X_i,\theta)/\sigma_{P,j}(\theta)) \xrightarrow{p} \begin{cases} 0 & \text{if } \ j \in \mathcal{J}^*(\theta) \\ -\infty & \text{otherwise,} \end{cases} \tag{G.14}$$

where the first of those cases uses nonpositivity of φ_j. It therefore suffices to show that the right hand side of G.13 strictly decreases in ρ if and only if the conditions of the Lemma hold.

To simplify notation, henceforth omit dependence of $\mathbb{G}_{P,j}(\theta)$, $D_P(\theta)$, and $\mathcal{J}^*(\theta)$ on P and θ. Define the J vector e to have elements $e_j = c - \mathbb{G}_j,\ j = 1,\dots,J$. Suppose for simplicity that $\mathcal{J}^*$ contains the first J^* inequality constraints. Let $e^{[1:J^*]}$ denote the subvector of e that only contains elements corresponding to $j \in \mathcal{J}^*$, define $D^{[1:J^*,:]}$ correspondingly, and write

$$K = \begin{bmatrix} D^{[1:J^*,:]} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix},\ g = \begin{bmatrix} e^{[1:J^*]} \\ \rho\cdot\mathbf{1}_d \\ \rho\cdot\mathbf{1}_d \\ 0 \\ 0 \end{bmatrix},\ \tau = \begin{bmatrix} 0\cdot\mathbf{1}_{J^*} \\ \mathbf{1}_d \\ \mathbf{1}_d \\ 0 \\ 0 \end{bmatrix}.$$

where I_d denotes the $d\times d$ identity matrix. By Farkas' Lemma (Rockafellar, 1970, Theorem 22.1), the linear system $K\lambda \leq g$ has a solution if and only if for all $\mu \in \mathbb{R}^{J^*+2d+2}_+$,

$$\mu'K = 0 \Rightarrow \mu'g \geq 0. \tag{G.15}$$

To further simplify expressions, fix $p = [1\ 0\ \dots\ 0]$. Let $\mathcal{M} = \{\mu \in \mathbb{R}^{J^*+2d+2}_+ : \mu'K = 0\}$.

Step 1. This step shows that

$$\begin{aligned} & P(\{\lambda \in \rho B^d_{n,\rho} : \mathbb{G}_{P,j} + D_{P,j}\lambda \leq c, j \in \mathcal{J}^*\} \cap \{p'\lambda = 0\} \neq \emptyset) \\ > \ & P(\{\lambda \in (\rho-\delta)\rho B^d_{n,\rho} : \mathbb{G}_{P,j} + D_{P,j}\lambda \leq c, j \in \mathcal{J}^*\} \cap \{p'\lambda = 0\} \neq \emptyset) \end{aligned} \tag{G.16}$$

if and only if the condition on gradients holds. This is done by showing that

$$P(\{\mu'g \geq 0\ \forall \mu \in \mathcal{M}\} \cap \{\mu'g - \delta\tau < 0\ \exists \mu \in \mathcal{M}\}) > 0. \tag{G.17}$$

under that same condition. The event $\{\mu'g \geq 0\ \forall\mu \in \mathcal{M}\}$ obtains if and only if

$$\min_{\mu \in \mathbb{R}^{J^*+2d+2}_+} \{\mu'g :\ \mu'K = 0\} \geq 0 \tag{G.18}$$

and analogously for $\mu'(g - \delta\tau) \geq 0$. The values of these programs are not affected by adding a constraint as follows:

$$\min_{\mu \in \mathbb{R}^{J^*+2d+2}_+} \left\{ \mu'g :\ \mu'K = 0,\ \mu \in \arg\min_{\tilde{\mu} \in \mathbb{R}^{J^*+2d+2}_+} (\tilde{\mu}'g :\ \tilde{\mu}^{[1:J^*]} = \mu^{[1:J^*]}, \tilde{\mu}'K = 0) \right\}, \tag{G.19}$$

That is, we can restrict attention to a concentrated out subset of vectors μ, where the last $(2d+2)$ components of any μ minimize the objective function among all vectors that agree with μ in the first J^* components. The inner minimization problem in equation (G.19) can be written as

$$\min_{\tilde{\mu}^{[J^*+1:J^*+2d+2]}\in\mathbb{R}_+^{2d+2}} \rho \sum_{j=J^*+1}^{J^*+2d} \tilde{\mu}_j \quad \text{s.t.} \begin{bmatrix} \tilde{\mu}_{J^*+1}-\tilde{\mu}_{J^*+d+1}+\tilde{\mu}_{J^*+2d+1}-\tilde{\mu}_{J^*+2d+2} \\ \tilde{\mu}_{J^*+2}-\tilde{\mu}_{J^*+d+2} \\ \vdots \\ \tilde{\mu}_{J^*+d}-\tilde{\mu}_{J^*+2d} \end{bmatrix} = -\mu^{[1:J^*]\prime} D^{[1:J^*,:]}. \quad \text{(G.20)}$$

Thus, the solution of the problem is uniquely pinned down as

$$\mu^{[J^*+1:J^*+2d+2]} = \begin{bmatrix} 0 \\ -\left[D^{[1:J^*,2:d]\prime}\mu^{[1:J^*]} \wedge 0\cdot \mathbf{1}_{d-1}\right] \\ 0 \\ D^{[1:J^*,2:d]\prime}\mu^{[1:J^*]} \vee 0\cdot \mathbf{1}_{d-1} \\ -\left[D^{[1:J^*,1]\prime}\mu^{[1:J^*]} \wedge 0\right] \\ D^{[1:J^*,1]\prime}\mu^{[1:J^*]} \vee 0 \end{bmatrix}, \quad \text{(G.21)}$$

where $D^{[1:J^*,2:d]\prime}\mu^{[1:J^*]} \vee 0\cdot\mathbf{1}_{d-1}$ indicates a component-wise comparison. Now we consider the following case distinction:

Case (i). If $D_j/\|D_j\| \in \{p,-p\}$ for all $j \in \mathcal{J}^*$, then $\mu^{[1:J^*]\prime}D = (\mu^{[1:J^*]\prime}D^{[1:J^*,1]}, 0, ..., 0)'$ and therefore all but the last two entries of $\mu^{[J^*+1:J^*+2d+2]}$ equal zero. One can, therefore, restrict attention to vectors μ with $\mu^{[J^*+1:J^*+2d]} = 0$. But for these vectors, $\mu'\tau = 0$ and so the programs we compare necessarily have the same value. The probability in equation (G.17) is therefore zero.

Case (ii). Suppose that at least one row of D, say its first row (though it can be one direction of an equality constraint), is not collinear with p, so that $\|D^{[1,2:d]}\| \neq 0$.

Let

$$\varpi = \begin{bmatrix} 1 \\ 0\cdot\mathbf{1}_{J^*-1} \\ 0 \\ -\left[(D^{[1,2:d]\prime}) \wedge 0\cdot\mathbf{1}_{d-1}\right] \\ 0 \\ (D^{[1,2:d]\prime}) \vee 0\cdot\mathbf{1}_{d-1} \\ -\left[(D^{[1,1]}) \wedge 0\right] \\ (D^{[1,1]}) \vee 0 \end{bmatrix} \quad \text{(G.22)}$$

and note that $\varpi^{[J^*+1:J^*+2d]} \neq 0$, hence $\varpi'\tau > 0$.

As in the proof of Lemma E.6, the set $\mathcal{M}$ can be expressed as positive span of a finite, nonstochastic set of affinely independent vectors $\nu^t \in \mathbb{R}_+^{J^*+2d+2}$ that are determined only up to multiplication by a positive scalar. All of these vectors have the "concentrated out structure" in equation (G.21). But then ϖ must be one of them because it is the unique concentrated out vector with $\varpi^{[1:J^*]} = (1,0,...,0)'$, and $(1,0,...,0)'$ cannot be spanned by nonnegative J^*-vectors other than positive multiples of itself.

We now establish positive probability of the event

$$\begin{aligned} \nu^{t\prime} g &\geq 0, \text{ all } \nu^t \\ \nu^{t\prime}(g-\delta\tau) &< 0, \text{ some } \nu^t \end{aligned}$$

by observing that if we define

$$\iota_k = \begin{bmatrix} -\rho \cdot \sum_{i=2}^{d} \left| D^{[1,i]} \right| \\ k \cdot \mathbf{1}_{J^*-1} \\ \rho \cdot \mathbf{1}_d \\ \rho \cdot \mathbf{1}_d \\ 0 \\ 0 \end{bmatrix}, \tag{G.23}$$

then we have

$$0 = \varpi' \iota_k = \min_t \nu^{t\prime} \iota_k.$$

Any other spanning vector ν^t will not have $\varpi^{[2:J^*]} = 0$ and so for any such vector, $\nu^{t\prime}\iota_k$ strictly increases in k. As there are finitely many spanning vectors, all of them have strictly positive inner product with ι_k if k is chosen large enough.

A realization of $g = \iota_k$ would, therefore, yield

$$\nu^{t\prime} g \geq 0 \ \forall \nu^t \in \mathcal{M}, \text{ and } \varpi^{t\prime} (g - \delta\tau) < -\epsilon, \tag{G.24}$$

for some $\epsilon > 0$. Let

$$\Gamma_k = \{\iota : \iota = \iota_k + \epsilon/2b, \ \|b\| \leq 1 \text{ and } \varpi' b > 0\}. \tag{G.25}$$

Then

$$\nu^{t\prime} \iota \geq 0 \ \forall \nu^t \in \mathcal{M}, \text{ and } \varpi^{t\prime} (\iota - \delta\tau) < -\epsilon/2, \ \forall \iota \in \Gamma_k. \tag{G.26}$$

The probability in equation (G.17) is therefore strictly positive.

Step 2. Next, we argue that

$$P(\{\lambda \in \rho B_{n,\rho}^d : \ \mathbb{G}_j + D_j \lambda \leq c, j \in \mathcal{J}^*\} \cap \{p'\lambda = 0\} \neq \emptyset) \tag{G.27}$$

strictly continuously increases in c. The rigorous argument is very similar to the use of Farkas' Lemma in step 1 and in Lemma E.6. We leave it at an intuition: As c increases, the set of vectors g fulfilling the right hand side of (G.15) strictly increases, hence the set of realizations of $\mathbb{G}_j$ that render the program feasible strictly increases, and $\mathbb{G}_j$ has full support.

Step 3. Steps 1 and 2 imply that

$$\begin{aligned} & \inf_{c \geq 0} \{P(\{\lambda \in \rho B_{n,\rho}^d : \mathbb{G}_j + D_j \lambda \leq c, j \in \mathcal{J}^*\} \cap \{p'\lambda = 0\} \neq \emptyset) \geq 1 - \alpha\} \\ > \ & \inf_{c \geq 0} \{P(\{\lambda \in (\rho - \delta)\rho B_{n,\rho}^d : \mathbb{G}_j + D_j \lambda \leq c, j \in \mathcal{J}^*\} \cap \{p'\lambda = 0\} \neq \emptyset) \geq 1 - \alpha\} \end{aligned} \tag{G.28}$$

and hence the result. ∎

References

Adams, R. A., and J. J. Fournier (2003): *Sobolev spaces*, vol. 140. Academic press.

Andrews, D. W. (1994): "Empirical process methods in econometrics," *Handbook of econometrics*, 4, 2247–2294.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-In Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25(3), 669–709.

——— (2010): "Asymptotic Size and a Problem With Subsampling and With the m Out Of n Bootstrap," *Econometric Theory*, 26, 426–468.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Beresteanu, A., and F. Molinari (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76, 763–814.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

Brent, R. P. (1971): "An algorithm with guaranteed convergence for finding a zero of a function," *The Computer Journal*, 14(4), 422–425.

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259–282.

——— (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8(1), 1–38.

Bull, A. D. (2011): "Convergence rates of efficient global optimization algorithms," *Journal of Machine Learning Research*, 12(Oct), 2879–2904.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets In Econometric Models," *Econometrica*, 75, 1243–1284.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Davydov, Y. A., M. Lifshitz, and N. Smorodina (1995): *Local properties of distributions of stochastic functionals*. American Mathematical Society.

Dekker, T. (1969): "Finding a zero by means of successive linear interpolation," *Constructive aspects of the fundamental theorem of algebra*, pp. 37–51.

Horn, R. A., and C. R. Johnson (1985): *Matrix Analysis*. Cambdridge University Press.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kaido, H., F. Molinari, and J. Stoye (2017): "Constraint qualifications in projection inference," Work in progress.

Mammen, E. (1992): *When Does Bootstrap Work? Asymptotic Results and Simulations*. Spinger Verlag, New York, NY.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

Molchanov, I. (2005): *Theory of Random Sets*. Springer, London.

Narcowich, F., J. Ward, and H. Wendland (2003): "Refined error estimates for radial basis function interpolation," *Constructive approximation*.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011): "Moment Inequalities and Their Application," Discussion Paper, Harvard University.

Pata, V. (2014): "Fixed Point Theorems and Applications," Mimeo.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton.

Rockafellar, R. T., and R. J.-B. Wets (2005): *Variational Analysis, Second Edition*. Springer-Verlag, Berlin.

Steinwart, I., and A. Christmann (2008): *Support vector machines*. Springer Science & Business Media.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

Tartar, L. (2007): *An introduction to Sobolev spaces and interpolation spaces*, vol. 3. Springer Science & Business Media.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, Berlin.

van der Vaart, A. W., and J. H. van Zanten (2008): "Reproducing kernel Hilbert spaces of Gaussian priors," in *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pp. 200–222. Institute of Mathematical Statistics.