*Econometrica Supplementary Material*

# ONLINE APPENDIX TO "CONFIDENCE INTERVALS FOR PROJECTIONS OF PARTIALLY IDENTIFIED PARAMETERS"

HIROAKI KAIDO[1], FRANCESCA MOLINARI[2] AND JÖRG STOYE[3]

## OUTLINE

Section D states and proofs Theorem D.1, which establishes convergence-related results for our E-A-M algorithm. It also 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 E.1 presents the assumptions under which we prove asymptotic uniform validity of coverage of our procedure. Section F verifies some of our main assumptions for moment (in)equality models that have received much attention in the literature. Section G summarizes the notation we use and the structure of the proof of Theorem 3.1,[1] and provides a proof of Theorems 3.1 (both under our main assumptions and under a high level assumption replacing Assumption E.3 and dropping the $\rho$-box constraints). Section H contains the statements and proofs of the lemmas used to establish Theorems 3.1 and D.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 3.1.

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

## D. ADDITIONAL CONVERGENCE RESULTS AND BACKGROUND MATERIALS FOR THE E-A-M ALGORITHM AND FOR COMPUTATION OF $\hat{c}_n(\theta)$

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

#### D. *Assumption D.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 (D.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 F.1 we verify the condition for the canonical examples in the moment (in)equality literature.

ASSUMPTION D.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)* $\varphi_j$ *is Lipschitz continuous in* $x \in \mathbb{R}$ *for all* $j = 1, \ldots, J$.

Department of Economics, Boston University, hkaido@bu.edu.
Department of Economics, Cornell University, fm72@cornell.edu
Department of Economics, Cornell University, stoye@cornell.edu

[1]Financial support from NSF grants SES-1230071, SES-1357643 and SES-1824344 is gratefully acknowledged.
[2]Financial support from NSF grants SES-0922330 and SES-1824375 is gratefully acknowledged.
[3]Financial support from NSF grants SES-1260980 and SES-1824375 is gratefully acknowledged.

[1]Section G.1 provides in Table G.1 a summary of the notation used throughout, and in Figure G.1 and Table G.2 a flow diagram and heuristic explanation of how each lemma contributes to the proof of Theorem 3.1.

*D. Statement and Proof of Theorem D.1*

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

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

where the family of functions $\phi_\tau$ 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 D.1 *Suppose Assumptions E.1, E.2, E.4, E.5 and D.1 hold. Let $\tau_n$ 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.*

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

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

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

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

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

*For $c \geq 0$, let $U_n(\theta_n, c)$ be defined as in (G.26). Then,*

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

*4. Fix $P \in \mathcal{P}$ and $n$. There exists $R > 0$ such that $\|\hat{c}_{n,\tau_n}\|_{\mathcal{H}_\beta} \leq R$.*

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

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

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 $\infty$. By Assumption D.1-(i) and arguing as in the proof of Theorem 2 in Andrews (1994), condition (H.224) in Lemma H.11 is satisfied with $v = d$. Also, by Lemma H.13, (H.225) in Lemma H.11 holds with $\gamma = 1$. This therefore ensures the conditions of Lemma H.11.

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

$$\Big|\frac{m_j^2(x,\theta)}{\sigma_{P,j}^2(\theta)}-\frac{m_j^2(x,\theta)}{\sigma_{P,j}^2(\theta)}\Big|\le\Big|\frac{m_j(x,\theta)}{\sigma_{P,j}(\theta)}+\frac{m_j(x,\theta')}{\sigma_{P,j}(\theta')}\Big|\Big|\frac{m_j(x,\theta)}{\sigma_{P,j}(\theta)}-\frac{m_j(x,\theta')}{\sigma_{P,j}(\theta')}\Big| \tag{D.7}$$

$$\le 2F(x)\bar{M}(x)\|\theta-\theta'\|. \tag{D.8}$$

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

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

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

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

Further, for any $C>0$

$$\begin{aligned} E_P[\bar{F}^2(X)1\{\bar{F}(X)>C\}] \;&\le\; E_P[\bar{F}^2(X)]P(\bar{F}(X)>C) \\ &\le\; 4E_P[|F(X)M(X)|^2]\frac{\|\bar{F}\|_{L_P^1}}{C}\;\le\;\frac{4M^2}{C}, \end{aligned} \tag{D.11}$$

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 H.12 (i). Note also that Assumption D.1-(i) ensures the conditions of Lemma H.12 (ii). Therefore, by Lemmas H.11-H.12 and Assumption E.4, for any $\eta>0$, there exists $C>0$ such that $\inf_{P\in\mathcal{P}}P(E_n)\ge 1-\eta$ for all $n$ sufficiently large.

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

$$\begin{aligned} &\Big|\mathbb{G}_{n,j}^b(\theta)+\rho\hat{D}_{n,j}(\theta)\lambda+\varphi_j(\hat{\xi}_{n,j}(\theta))-\mathbb{G}_{n,j}^b(\theta')-\rho\hat{D}_{n,j}(\theta')\lambda-\varphi_j(\hat{\xi}_{n,j}(\theta'))\Big| \\ &\le|\mathbb{G}_{n,j}^b(\theta)-\mathbb{G}_{n,j}^b(\theta')|+\rho\|\hat{D}_{n,j}(\theta)-\hat{D}_{n,j}(\theta')\|\sup_{\lambda\in B^d}\|\lambda\|+|\varphi_j(\hat{\xi}_{n,j}(\theta))-\varphi_j(\hat{\xi}_{n,j}(\theta'))|. \end{aligned} \tag{D.12}$$

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 $\mathbb{G}_{n,j}^b(\theta)-\mathfrak{G}_{n,j}^b(\theta)=\mathfrak{G}_{n,j}^b(\theta)\eta_{n,j}(\theta)$,

$$\begin{aligned} |\mathbb{G}_{n,j}^b(\theta)-\mathbb{G}_{n,j}^b(\theta')|&\le|\mathfrak{G}_{n,j}^b(\theta)-\mathfrak{G}_{n,j}^b(\theta')|+2\sup_{\theta\in\Theta}|\mathfrak{G}_{n,j}^b(\theta)|\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \\ &\le|\mathfrak{G}_{n,j}^b(\theta)-\mathfrak{G}_{n,j}^b(\theta')|+2\sup_{\theta\in\Theta}|\mathfrak{G}_{n,j}^b(\theta)|\frac{C}{\sqrt{n}}. \end{aligned} \tag{D.13}$$

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

$$F_n\equiv\big\{m_n\in Q:\sup_{\|\theta-\theta'\|\le\tau_n}\|\mathfrak{G}_n^b(\theta)-\mathfrak{G}_n^b(\theta')\|\le(\ln n)^2\tau_n,\ \sup_{\theta\in\Theta}\|\mathfrak{G}_n^b(\theta)\|\le C\big\}. \tag{D.14}$$

By Lemma H.11 (ii) and the asymptotic tightness of $\mathfrak{G}_n^b$, for any $\eta>0$, there exists a $C$ such that $P_n^*(F_n)\ge 1-\eta$ 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 (D.13) is bounded by $(\ln n)^2\tau_n + C/\sqrt{n}$ for some $C>0$.

Next, by the triangle inequality and Assumption E.4,

$$\begin{aligned}(D.15)\quad & \|\hat{D}_{n,j}(\theta)-\hat{D}_{n,j}(\theta')\| \\ & \le \|\hat{D}_{n,j}(\theta)-D_{P,j}(\theta)\| + \|D_{P,j}(\theta)-D_{P,j}(\theta')\| + \|\hat{D}_{n,j}(\theta')-D_{P,j}(\theta')\| \le C\beta_n + C\tau_n.\end{aligned}$$

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

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

Hence,

$$\begin{aligned}(D.17)\quad & |\hat{\xi}_{n,j}(\theta)-\hat{\xi}_{n,j}(\theta')| \le \kappa_n^{-1}|\mathbb{G}_{n,j}(\theta)-\mathbb{G}_{n,j}(\theta')| \\ &\quad + \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta)}{\sigma_{P,j}(\theta)}\eta_{n,j}(\theta) - \frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\eta_{n,j}(\theta')\Big| + \kappa_n^{-1}\sqrt{n}D_{P,j}(\bar{\theta})\|\theta-\theta'\|.\end{aligned}$$

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

$$\begin{aligned}(D.18)\quad & \kappa_n^{-1}(\ln n)^2\tau_n + \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - \frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\Big||\eta_{n,j}(\theta)| \\ &\quad + \kappa_n^{-1}\sqrt{n}\Big|\frac{\bar{m}_{n,j}(\theta')}{\sigma_{P,j}(\theta')}\Big||\eta_{n,j}(\theta)-\eta_{n,j}(\theta')| + C\kappa_n^{-1}\sqrt{n}\tau_n \\ &\qquad \le \kappa_n^{-1}(\ln n)^2\tau_n + \kappa_n^{-1}\sqrt{n}\tau_n\frac{C}{\sqrt{n}} + C\kappa_n^{-1}\sqrt{n}\tau_n + C\kappa_n^{-1}\sqrt{n}\tau_n,\end{aligned}$$

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

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

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

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

$$\begin{aligned}(D.20)\quad & \mathbb{G}^b_{n,j}(\theta') + \hat{D}_{n,j}(\theta')\lambda + \varphi_j(\hat{\xi}_{n,j}(\theta')) \\ &\qquad \le \mathbb{G}^b_{n,j}(\theta) + \hat{D}_{n,j}(\theta)\lambda + \varphi_j(\hat{\xi}_{n,j}(\theta)) + C\varepsilon_n \le \hat{c}_n(\theta) + C\varepsilon_n,\end{aligned}$$

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

$$
\begin{aligned}
\text{(D.21)}\quad & P_n^*\left(\left\{\Lambda_n^b(\theta', \rho, \hat{c}_n(\theta)+C\varepsilon_n)\cap\{p'\lambda=0\}\neq\emptyset\right\}\right)\\
&\geq P_n^*\left(\left\{\Lambda_n^b(\theta', \rho, \hat{c}_n(\theta)+C\varepsilon_n)\cap\{p'\lambda=0\}\neq\emptyset\right\}\cap F_n\right)\\
&\qquad\qquad \geq P_n^*\left(\left\{\Lambda_n^b(\theta, \rho, \hat{c}_n(\theta))\cap\{p'\lambda=0\}\neq\emptyset\right\}\cap F_n\right)\geq 1-\alpha-\eta.
\end{aligned}
$$

Since $\eta$ is arbitrary, we have

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

Reversing the roles of $\theta$ and $\theta'$ 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 (D.3), we use that for any $\theta, \theta' \in \Theta$ such that $\|\theta-\theta'\| \leq \tau_n$, $|\hat{c}_n(\theta) - \hat{c}_n(\theta')| \leq C\varepsilon_n$ with probability approaching 1 uniformly in $P \in \mathcal{P}$ by the result in Part 1. This implies

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

**Part 3.** By Part 2 and the definition of $\hat{c}_{n,\rho,\tau}$ in (D.4), it follows that

$$
\begin{aligned}
\text{(D.22)}\quad \hat{c}_{n,\rho,\tau_n}(\theta_n) &\geq \hat{c}_{n,\rho}(\theta_n) - e_n\\
&\geq c^I_{n,\rho}(\theta_n) - e_n,
\end{aligned}
$$

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 H.1. Note that Lemma H.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 (H.22) with $\hat{c}_{n,\rho,\tau_n}$ replacing $\hat{c}_{n,\rho}$, and mimic the argument following (H.22) in the proof of Lemma H.1. Then, the conclusion of the lemma follows.

**Part 4.** 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 $|\alpha| \leq s$, the partial derivative $\nabla^\alpha \hat{c}_{n,\tau_n}$ is bounded by some constant $M > 0$ on the compact set $\Theta$, and hence

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

where $\upsilon$ denote the Lebesgue measure on $\mathbb{R}^d$. This ensures $\nabla^\alpha \hat{c}_{n,\tau_n} \in L^2_\upsilon(\Theta)$ for all $|\alpha| \leq s$. 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 $\|\hat{c}_{n,\tau_n}\|_{\mathcal{H}_\beta} \leq C\|\hat{c}_{n,\tau_n}\|_{H^{\nu+d/2}} \leq R$.

*Q.E.D.*

*D. 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

$$K_\beta(\theta-\theta')=\exp\big(-\sum_{k=1}^{d}|(\theta_k-\theta'_k)/\beta_k|^2\big),\ \beta_k\in[\underline{\beta}_k,\overline{\beta}_k],\ k=1,\cdots,d, \tag{D.23}$$

where $0<\underline{\beta}_k<\overline{\beta}_k<\infty$ for all $k$. The second one is the class of Matérn kernels (see, e.g., Rasmussen and Williams, 2005, Chapter 4) 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 $k_\nu$ is the modified Bessel function of the second kind.[2] The index $\nu$ controls the smoothness of $K_\beta$. 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

$$\|f\|_{\mathcal{H}_\beta}\equiv\inf_{g|_D=f}\int\frac{\hat{g}(\zeta)}{\hat{K}_\beta(\zeta)}d\zeta, \tag{D.24}$$

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

$$\|f\|^2_{H^{\nu+d/2}}\equiv\inf_{g|_{D^o}=f}\int\frac{\hat{g}(\zeta)}{(1+\|\zeta\|^2)^{\nu+d/2}}d\zeta \tag{D.25}$$

[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 $2\nu$-Hölder continuity of $K_\beta$ at the origin without a log factor (Bull, 2011, Assumption 4).

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).

Theorem 3.2 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 D.1 by introducing a mollified version of $\hat{c}_n$.

### D. *A Reformulation of the M-step as a Nonlinear Program*

In (2.21), $\theta^{(L+1)}$ is defined as the maximizer of the following maximization problem

$$\max_{\theta\in\Theta}(p'\theta - p'\theta^*_L)_+\Big(1-\Phi\Big(\frac{\bar{g}(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big), \tag{D.26}$$

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

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

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

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

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

$$\max_{(\theta,v)\in\Theta\times\mathbb{R}} v \ \text{ s.t. } (p'\theta - p'\theta^*_L)_+\Big(1-\Phi\Big(\frac{g_j(\theta)-c_L(\theta)}{\hat{\varsigma}s_L(\theta)}\Big)\Big) \geq v,\ j=1,\dots,J,$$

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.

### D. *Root-Finding Algorithm Used to Compute $\hat{c}_n(\theta)$*

This section explains in detail how $\hat{c}_n(\theta)$ in equation (2.13) is computed. For a given $\theta \in \Theta$, $P^*(\Lambda^b_n(\theta,\rho,c)\cap\{p'\lambda=0\}\neq\emptyset)$ increases in $c$ (with $\Lambda^b_n(\theta,\rho,c)$ defined in (2.11)), 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

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

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

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

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

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

(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 $h_\alpha(c_U) \leq Tol$ or if $|c_U - c_L| \leq Tol$, 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.

## E. Assumptions for Asymptotic Coverage Validity

### *E. Main Assumptions*

We posit that $P$, the distribution of the observed data, belongs to a class of distributions denoted by $\mathcal{P}$. We write stochastic order relations that hold uniformly over $P \in \mathcal{P}$ using the notations $o_{\mathcal{P}}$ and $O_{\mathcal{P}}$; see Appendix G.1 for the formal definitions. Below, $\epsilon$, $\varepsilon$, $\delta$, $\omega$, $\underline{\sigma}$, $M$, $\bar{M}$ denote generic

[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.

constants which may be different in different appearances but cannot depend on $P$. Given a square matrix $A$, we write $\mathrm{eig}(A)$ for its smallest eigenvalue.

ASSUMPTION E.1 *(a) $\Theta \subset \mathbb{R}^d$ is a compact hyperrectangle with nonempty interior.*
*(b) All distributions $P \in \mathcal{P}$ satisfy the following:*
*(i) $E_P[m_j(X_i,\theta)] \leq 0,\ j = 1,\dots,J_1$ and $E_P[m_j(X_i,\theta)] = 0,\ j = J_1+1,\dots,J_1+J_2$ for some $\theta \in \Theta$;*
*(ii) $\{X_i, i \geq 1\}$ are i.i.d.;*
*(iii) $\sigma^2_{P,j}(\theta) \in (0,\infty)$ for $j = 1,\dots,J$ for all $\theta \in \Theta$;*
*(iv) For some $\delta > 0$ and $M \in (0,\infty)$ and for all $j$, $E_P[\sup_{\theta\in\Theta} |m_j(X_i,\theta)/\sigma_{P,j}(\theta)|^{2+\delta}] \leq M$.*

ASSUMPTION E.2 *The function $\varphi_j$ is continuous at all $x \geq 0$ and $\varphi_j(0) = 0$; $\kappa_n \to \infty$ and $\kappa_n = o(n^{1/2})$. If Assumption E.3-2 is imposed, $\kappa_n = o(n^{1/4})$.*

Assumption E.1-(a) requires that $\Theta$ is a hyperrectangle, but can be replaced with the assumption that $\theta$ is defined through a finite number of nonstochastic inequality constraints smooth in $\theta$ and such that $\Theta$ is convex. Compactness is a standard assumption on $\Theta$ for extremum estimation. We additionally require convexity as we use mean value expansions of $E_P[m_j(X_i,\theta)]/\sigma_{P,j}(\theta)$ in $\theta$; see (2.9). Assumption E.1-(b) defines our moment (in)equalities model. Assumption E.2 constrains the GMS function and the rate at which its tuning parameter diverges. Both E.1-(b) and E.2 are based on Andrews and Soares (2010) and are standard in the literature,[4] although typically with $\kappa_n = o(n^{1/2})$. The slower rate $\kappa_n = o(n^{1/4})$ is satisfied for the popular choice, recommended by Andrews and Soares (2010), of $\kappa_n = \sqrt{\ln n}$.

Next, and unlike some other papers in the literature, we impose restrictions on the correlation matrix of the moment functions. These conditions can be easily verified in practice because they are implied when the correlation matrix of the moment equality functions and the moment inequality functions specified below have a determinant larger than a predefined constant for any $\theta \in \Theta$.

ASSUMPTION E.3 *All distributions $P \in \mathcal{P}$ satisfy **one** of the following two conditions for some constants $\omega > 0, \underline{\sigma} > 0, \epsilon > 0, \varepsilon > 0, M < \infty$:*

1. *Let $\mathcal{J}(P,\theta;\varepsilon) \equiv \{j \in \{1,\cdots,J_1\} : E_P[m_j(X_i,\theta)]/\sigma_{P,j}(\theta) \geq -\varepsilon\}$. Denote*

$$\tilde{m}(X_i,\theta) \equiv \left(\{m_j(X_i,\theta)\}_{j\in\mathcal{J}(P,\theta;\varepsilon)}, m_{J_1+1}(X_i,\theta),\dots,m_{J_1+J_2}(X_i,\theta)\right)',$$
$$\tilde{\Omega}_P(\theta) \equiv Corr_P(\tilde{m}(X_i,\theta)).$$

   *Then $\inf_{\theta\in\Theta_I(P)} \mathrm{eig}(\tilde{\Omega}_P(\theta)) \geq \omega$.*

2. *The functions $m_j(X_i,\theta)$ are defined on $\Theta^\epsilon = \{\theta \in \mathbb{R}^d : d(\theta,\Theta) \leq \epsilon\}$. There exists $R_1 \in \mathbb{N}$, $1 \leq R_1 \leq J_1/2$, and measurable functions $t_j : \mathcal{X} \times \Theta^\epsilon \to [0,M]$, $j \in \mathcal{R}_1 \equiv \{1,\dots,R_1\}$, such that for each $j \in \mathcal{R}_1$,*

$$m_{j+R_1}(X_i,\theta) = -m_j(X_i,\theta) - t_j(X_i,\theta). \tag{E.1}$$

[4] Continuity of $\varphi_j$ for $x \geq 0$ is restrictive only for GMS function $\varphi^{(2)}$ in Andrews and Soares (2010).

*For each* $j \in \mathcal{R}_1 \cap \mathcal{J}(P,\theta;\varepsilon)$ *and any choice* $\ddot{m}_j(X_i,\theta) \in \{m_j(X_i,\theta), m_{j+R_1}(X_i,\theta)\}$*, denoting* $\tilde{\Omega}_P(\theta) \equiv Corr_P(\tilde{m}(X_i,\theta))$*, where*

$$\tilde{m}(X_i,\theta) \equiv \Big(\{\ddot{m}_j(X_i,\theta)\}_{j\in\mathcal{R}_1\cap\mathcal{J}(P,\theta;\varepsilon)},\\ \{m_j(X_i,\theta)\}_{j\in\mathcal{J}(P,\theta;\varepsilon)\setminus\{1,\dots,2R_1\}}, m_{J_1+1}(X_i,\theta),\dots,m_{J_1+J_2}(X_i,\theta)\Big)',$$

*one has*

$$\inf_{\theta\in\Theta_I(P)} \mathrm{eig}(\tilde{\Omega}_P(\theta)) \geq \omega. \tag{E.2}$$

*Finally,*

$$\inf_{\theta\in\Theta_I(P)} \sigma_{P,j}(\theta) > \underline{\sigma} \text{ for } j = 1,\dots,R_1. \tag{E.3}$$

Assumption E.3-1 requires that the correlation matrix of the moment functions corresponding to close-to-binding moment conditions has eigenvalues uniformly bounded from below. This assumption holds in many applications of interest, including: (i) instances when the data is collected by intervals with minimum width;[5] (ii) in treatment effect models with (uniform) overlap; (iii) in static complete information entry games under weak solution concepts, e.g. rationality of level 1, see Aradillas-Lopez and Tamer (2008).

We are aware of two examples in which Assumption E.3-1 may fail. One are missing data scenarios, e.g. scalar mean, linear regression, and best linear prediction, with a vanishing probability of missing data. The other example, which is extensively simulated in Section C, is the Ciliberto and Tamer (2009) entry game model when the solution concept is pure strategy Nash equilibrium. We show in Appendix F.2 that these examples satisfy Assumption E.3-2.

REMARK E.1 *Assumption E.3-2 weakens E.3-1 by allowing for (drifting to) perfect correlation among moment inequalities that cannot cross. This assumption is often satisfied in moment conditions that are separable in data and parameters, i.e. for each* $j = 1,\dots,J$,

$$E_P[m_j(X_i,\theta)] = E_P[h_j(X_i)] - v_j(\theta), \tag{E.4}$$

*for some measurable functions* $h_j : \mathcal{X} \to \mathbb{R}$ *and* $v_j : \Theta \to \mathbb{R}$*. Models like the one in Ciliberto and Tamer (2009) fall in this category, and we verify Assumption E.3-2 for them in Appendix F.2. The argument can be generalized to other separable models.*

*In Appendix F.2, we also verify Assumption E.3-2 for some models that are not separable in the sense of equation (E.4), for example best linear prediction with interval outcome data. The proof can be extended to cover (again non-separable) binary models with discrete or interval valued covariates under the assumptions of Magnac and Maurin (2008).*

[5] Empirically relevant examples are that of: (a) the Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics, which collects wage data from employers as intervals of positive width, and uses these data to construct estimates for wage and salary workers in 22 major occupational groups and 801 detailed occupations; and (b) when, due to concerns for privacy, data is reported as the number of individuals who belong to each of a finite number of cells (for example, in public use tax data).

In what follows, we refer to pairs of inequality constraints indexed by $\{j, j+R_1\}$ and satisfying (E.1) as "paired inequalities." Their presence requires a modification of the bootstrap procedure. This modification exclusively concerns the definition of $\Lambda_n^b(\theta, \rho, c)$ in equation (2.11). We explain it here for the case that the GMS function $\varphi_j$ is the hard-thresholding one in footnote 8 of the main paper, and refer to Appendix H equations (H.12)-(H.13) for the general case. If

$$\varphi_j(\hat{\xi}_{n,j}(\theta)) = 0 = \varphi_j(\hat{\xi}_{n,j+R_1}(\theta)),$$

we replace $\mathbb{G}_{n,j+R_1}^b(\theta)$ with $-\mathbb{G}_{n,j}^b(\theta)$ and $\hat{D}_{n,j+R_1}(\theta)$ with $-\hat{D}_{n,j}(\theta)$, so that inequality $\mathbb{G}_{n,j+R_1}^b(\theta) + \hat{D}_{n,j+R_1}(\theta)\lambda \leq c$ is replaced with $-\mathbb{G}_{n,j}^b(\theta) - \hat{D}_{n,j}(\theta)\lambda \leq c$ in equation (2.11). In words, when hard threshold GMS indicates that both paired inequalities bind, we pick one of them, treat it as an equality, and drop the other one. In the proof of Theorem 3.1, we show that this tightens the stochastic program.[6] The rest of the procedure is unchanged.

Instead of Assumption E.3, BCS (Assumption 2) impose the following high-level condition: (a) The limit distribution of their profiled test statistic is continuous at its $1-\alpha$ quantile if this quantile is positive; (b) else, their test is asymptotically valid with a critical value of zero. In Appendix G.2.2, we show that we can replace Assumption E.3 with a weaker high level condition (Assumption E.6) that resembles the BCS assumption but constrains the limiting coverage probability. (We do not claim that the conditions are equivalent.) The substantial amount of work required for us to show that Assumption E.3 implies Assumption E.6 is suggestive of how difficult these high-level conditions can be to verify.[7] Moreover, in Appendix E.3 we provide a simple example that violates Assumption E.3 and in which all of calibrated projection, BCS-profiling, and the bootstrap procedure in Pakes, Porter, Ho, and Ishii (2011) fail. The example leverages the fact that when binding constraints are near-perfectly correlated, the projection may be estimated superconsistently, invalidating the simple nonparametric bootstrap.[8]

Together with imposition of the $\rho$-box constraints, Assumption E.3 allows us to dispense with restrictions on the local geometry of the set $\Theta_I(P)$. Restrictions of this type, which are akin to constraint qualification conditions, are imposed by BCS (Assumption A.3-(a)), Pakes, Porter, Ho, and Ishii (2011, Assumptions A.3-A.4), Chernozhukov, Hong, and Tamer (2007, Condition C.2), and elsewhere. In practice, they can be hard to verify or pre-test for. We study this matter in detail in Kaido, Molinari, and Stoye (2019).

We next lay out regularity conditions on the gradients of the moments.

Assumption E.4 *All distributions $P \in \mathcal{P}$ satisfy the following conditions:*

*(i) For each $j$, there exist $D_{P,j}(\cdot) \equiv \nabla_\theta \{E_P[m_j(X,\cdot)]/\sigma_{P,j}(\cdot)\}$ and its estimator $\hat{D}_{n,j}(\cdot)$ such that $\sup_{\theta \in \Theta^\epsilon} \|\hat{D}_{n,j}(\theta) - D_{P,j}(\theta)\| = o_{\mathcal{P}}(1)$.*

*(ii) There exist $M, \bar{M} < \infty$ such that for all $\theta, \tilde{\theta} \in \Theta^\epsilon$ $\max_{j=1,\dots,J} \|D_{P,j}(\theta) - D_{P,j}(\tilde{\theta})\| \leq M\|\theta - \tilde{\theta}\|$ and $\max_{j=1,\dots,J} \sup_{\theta \in \Theta_I(P)} \|D_{P,j}(\theta)\| \leq \bar{M}$.*

Assumption E.4 requires that each of the $J$ normalized population moments is differentiable, that its derivative is Lipschitz continuous, and that this derivative can be consistently estimated

[6] When paired inequalities are present, in equation (2.6) instead of $\hat{\sigma}_{n,j}$ we use the estimator $\hat{\sigma}_{n,j}^M$ specified in (H.196) in Lemma H.10 p.54 of the Appendix for $\sigma_{P,j}, j = 1, \dots, 2R_1$ (with $R_1 \leq J_1/2$ defined in the assumption). In equation (2.10) we use $\hat{\sigma}_{n,j}$ for all $j = 1, \dots, J$. To ease notation, we do not distinguish the two unless it is needed.

[7] Assumption E.3 is used exclusively to obtain the conclusions of Lemma H.6, H.7 and H.8, hence any alternative assumption that delivers such results can be used.

[8] The example we provide satisfies all assumptions explicitly stated in Pakes, Porter, Ho, and Ishii (2011), illustrating an oversight in their Theorem 2.

uniformly in $\theta$ and $P$.[9] We require these conditions because we use a linear expansion of the population moments to obtain a first-order approximation to the nonlinear programs defining $CI_n$, and because our bootstrap procedure requires an estimator of $D_P$.

A final set of assumptions is on the normalized empirical process. For this, define the variance semimetric $\varrho_P$ by

$$\varrho_P(\theta,\tilde{\theta}) \equiv \left\| \left\{ \left[ Var_P\left(\sigma_{P,j}^{-1}(\theta)m_j(X,\theta) - \sigma_{P,j}^{-1}(\tilde{\theta})m_j(X,\tilde{\theta})\right)\right]^{1/2} \right\}_{j=1}^{J} \right\|. \tag{E.5}$$

For each $\theta,\tilde{\theta} \in \Theta$ and $P$, let $Q_P(\theta,\tilde{\theta})$ denote a $J$-by-$J$ matrix whose $(j,k)$-th element is the covariance between $m_j(X_i,\theta)/\sigma_{P,j}(\theta)$ and $m_k(X_i,\tilde{\theta}))/\sigma_{P,k}(\tilde{\theta})$.

Assumption E.5 *All distributions $P \in \mathcal{P}$ satisfy the following conditions:*

*(i) The class of functions $\{\sigma_{P,j}^{-1}(\theta)m_j(\cdot,\theta) : \mathcal{X} \to \mathbb{R}, \theta \in \Theta\}$ is measurable for each $j = 1,\dots,J$.*

*(ii) The empirical process $\mathbb{G}_n$ with $j$-th component $\mathbb{G}_{n,j}$ is uniformly asymptotically $\varrho_P$-equicontinuous. That is, for any $\epsilon > 0$,*

$$\lim_{\delta \downarrow 0} \lim_{n\to\infty} \sup_{P \in \mathcal{P}} P\left( \sup_{\varrho_P(\theta,\tilde{\theta})<\delta} \|\mathbb{G}_n(\theta) - \mathbb{G}_n(\tilde{\theta})\| > \epsilon \right) = 0. \tag{E.6}$$

*(iii) $Q_P$ satisfies*

$$\lim_{\delta \downarrow 0} \sup_{\|(\theta_1,\tilde{\theta}_1)-(\theta_2,\tilde{\theta}_2)\|<\delta} \sup_{P \in \mathcal{P}} \|Q_P(\theta_1,\tilde{\theta}_1) - Q_P(\theta_2,\tilde{\theta}_2)\| = 0. \tag{E.7}$$

Under this assumption, the class of normalized moment functions is uniformly Donsker (Bugni, Canay, and Shi, 2015a). We use this fact to show validity of our method.

### E. *High Level Conditions Replacing Assumption E.3 and the $\rho$-Box Constraints*

Next, we consider two high level assumptions. The first one aims at informally mimicking Assumption A.2 in Bugni, Canay, and Shi (2017) and replaces Assumption E.3. The second one replaces the use of the $\rho$-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 E.6 *Consider any sequence $\{P_n,\theta_n\} \in \{(P,\theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ such that*

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

*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 $\Omega$. Let*

$$\mathfrak{w}_j(\lambda) \equiv \mathbb{Z}_j + \rho D_j \lambda + \pi_{1,j}^*. \tag{E.8}$$

[9]The requirements are imposed on $\Theta^\epsilon$. Under Assumption E.3-1 it suffices they hold on $\Theta$.

*Let*

$$\mathfrak{W}(c) \equiv \left\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \le c,\ \forall j = 1, \ldots, J\right\}, \tag{E.9}$$

$$c_{\pi^*} \equiv \inf\{c \in \mathbb{R}_+ : \Pr(\mathfrak{W}(c) \neq \emptyset) \ge 1 - \alpha\}. \tag{E.10}$$

*Then:*

1. *If* $c_{\pi^*} > 0$, $\Pr\left(\mathfrak{W}(c) \neq \emptyset\right)$ *is continuous and strictly increasing at* $c = c_{\pi^*}$*.*
2. *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* (G.26).

Assumption E.7 *Consider any sequence* $\{P_n, \theta_n\} \in \{(P, \theta) : P \in \mathcal{P}, \theta \in \Theta_I(P)\}$ *as in Assumption E.6. Let*

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

*which differs from* (E.9) *by not constraining* $\lambda$ *to* $\mathfrak{B}^d_\rho$*, 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.* $\Pr(\|\bar{\mathfrak{W}}(\bar{c})\|_H > M_\eta) \le \eta$*.*

## E. *Example of Methods Failure When Assumption E.3 Fails*

Consider one-sided testing with two inequality constraints in $\mathbb{R}^2$. The constraints are

$$\begin{aligned} \theta_1 + \theta_2 &\le E_P(X_1) \\ \theta_1 - \theta_2 &\le E_P(X_2). \end{aligned}$$

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 $\theta_1^*$ 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 E.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 $\theta$ 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.

## F. Verification of Assumptions for the Canonical Partial Identification Examples

In this section we verify: (i) Assumption D.1 which is the crucial condition in Theorem D.1, and (ii) Assumption E.3-2, for the canonical examples in the partial identification literature:

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 $P(W_0 \le W_1) = 1$. The identified set is defined as

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

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

$$\begin{aligned}\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R}^d : E_P(W_0|Z = z^r) - z^{r\prime}\theta \le 0,\\ z^{r\prime}\theta - E_P(W_1|Z = z^r) \le 0, r = 1, \dots, k\}.\end{aligned} \tag{F.2}$$

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

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

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

$$\Theta_I(P) = \{\theta \in \Theta \subset \mathbb{R}^d : \text{ equations (C.1), (C.2), (C.3), (C.4) hold for all } Z = z^r, \; r = 1, \ldots, k\}. \tag{F.4}$$

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

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

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

as it is done in, e.g., Ciliberto and Tamer (2009). We assume that for each of the four canonical examples under consideration, Assumption E.1 as well as one of the assumptions below hold.

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

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

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

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

## F. *Verification of Assumptions D.1 and A.2-(i)*

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 $n^{-1/2}$. The same arguments, with small modification, deliver verification of Assumption A.2-(i) provided $\hat{\sigma}_{n,j}(\theta)>0$.

1. **Mean with interval data**. Here $\sigma_{P,\ell}(\theta)=\sigma_{P,\ell}$, and under Assumption F.1 it is uniformly bounded from below. Then
$$\left|\frac{m_j(x,\theta)}{\sigma_{P,j}}-\frac{m_j(x,\theta')}{\sigma_{P,j}}\right|=\frac{\|(\theta'-\theta)\|}{\sigma_{P,j}},\quad \ell=0,1,$$
$$D_{P,\ell}(\theta)=\frac{(-1)^{(1-\ell)}}{\sigma_{P,\ell}},\quad \ell=0,1.$$
Assumption F.1 then guarantees that Assumption D.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 F.1-F.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 D.1 is satisfied for these rescaled functions by Assumptions F.1-F.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 H.12. Theorem D.1 then holds observing that $|P(Z=z^r)/(\sum_{i=1}^n\mathbf{1}(Z_i=z^r)/n)-1|=O_{\mathcal{P}}(n^{-1/2})$ and treating this random element similarly to how we treat $\eta_{n,j}(\cdot)$ in the proof of Theorem D.1.
3. **Best linear prediction with interval outcome data and discrete regressors.** Here
$$\text{(F.7)}\qquad m_r(X_i,\theta)=q^r[Z_i(Z_i'\theta-(W_{0,i}+\mathbf{1}(q^rZ_i>0)(W_{1,i}-W_{0,i})))]$$
hence is Lipschitz in $\theta$ with constant $Z_iZ_i'$. Under Assumptions F.1-F.2, $Var_P(m_r(X_i,\theta))$ is uniformly bounded from below, and Lipschitz in $\theta$ with a constant that depends on $Z_i^4$. Hence $\frac{m_r(X_i,\theta)}{\sigma_{P,r}(\theta)}$ is Lipschitz in $\theta$ with a constant that depends on powers of $Z$. Because $Z$ has bounded support, Assumption D.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 E.1 and F.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
$$\frac{\mathbf{1}(Y_1=s,Y_2=t,Z=z^r)/P(Z=z^r)-g_{str}(\theta)}{\sigma_{P,str}},\ (s,t)\in\{0,1\}\times\{0,1\},r=1,\dots,k,$$

and the gradient is

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

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.

## F. *Verification of Assumption E.3-2*

Here we verify Assumption E.3-2 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 (E.1)-(E.2) are satisfied by construction, equation (E.3) is directly assumed.
2. **Linear regression with interval outcome data and discrete regressors.** Equation (E.1) is satisfied by construction. Given the estimator that we use for the population moment conditions, we verify equation (E.3) for the variances of the limit distribution of the vector $[\sqrt{n}(\hat{E}_n(W_\ell|Z = z^r) - E_P(W_\ell|Z = z^r))]_{\ell \in \{0,1\}, r=1,\ldots,k}$. We then have that equation (E.3) follows from Assumption F.1. Concerning equation (E.3), this needs to be verified for the correlation matrix of the limit distribution of a $r \times 1$ random vector that for each $r = 1, \ldots, k$ 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 (E.2) holds because the correlation matrix is diagonal.
3. **Best linear prediction with interval outcome data and discrete regressors.** Equation (E.1) is again satisfied by construction. Equation (E.2) holds under Assumptions F.1-F.2. Equation (E.3) is verified to hold under Assumption F.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 (C.3) and (C.4) are paired, but the corresponding moment functions differ by the model implied probability of the region of multiplicity, hence equation (E.1) is satisfied by construction. Given the estimator that we use for the population moment conditions, we verify equations (E.2) and (E.3) for the variances and for the correlation matrix of the limit distribution of the vector $\sqrt{n}(\hat{E}_n(Y_1 = s, Y_2 = t|Z = z^r) - E_P(Y_1 = s, Y_2 = t|Z = z^r)_{(s,t) \in \{0,1\} \times \{0,1\}, r=1,\ldots,k})$, which suffices for our results to hold. Equation (E.2) holds provided that $|Corr(Y_{i1}(1 - Y_{i2}), Y_{i1}Y_{i2})| < 1 - \epsilon$ for some $\epsilon > 0$ and Assumption F.3 holds.[10] To see that equation (E.3) also holds, note that Assumption F.3 yields that $P(Y_{i1} = 1, Y_{i2} = 0, Z_i = z^r)$ is uniformly bounded away from 0 and 1, thereby implying that for each $(s,t) \in \{0,1\} \times \{0,1\}, r = 1, \ldots, k$, $(P(Y_1 = s, Y_2 = t|Z = z^r)(1 - P(Y_1 = s, Y_2 = t|Z = z^r)))/(P(Z = z^r)(1 - P(Z = z^r)))$ is uniformly bounded away from zero.

[10] 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.

# G. Proof of Theorem 3.1

## G. Notation and Structure of the Proof of Theorem 3.1

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 G.1

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 (G.3)-(G.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 $\lambda \in \mathbb{R}^d$.

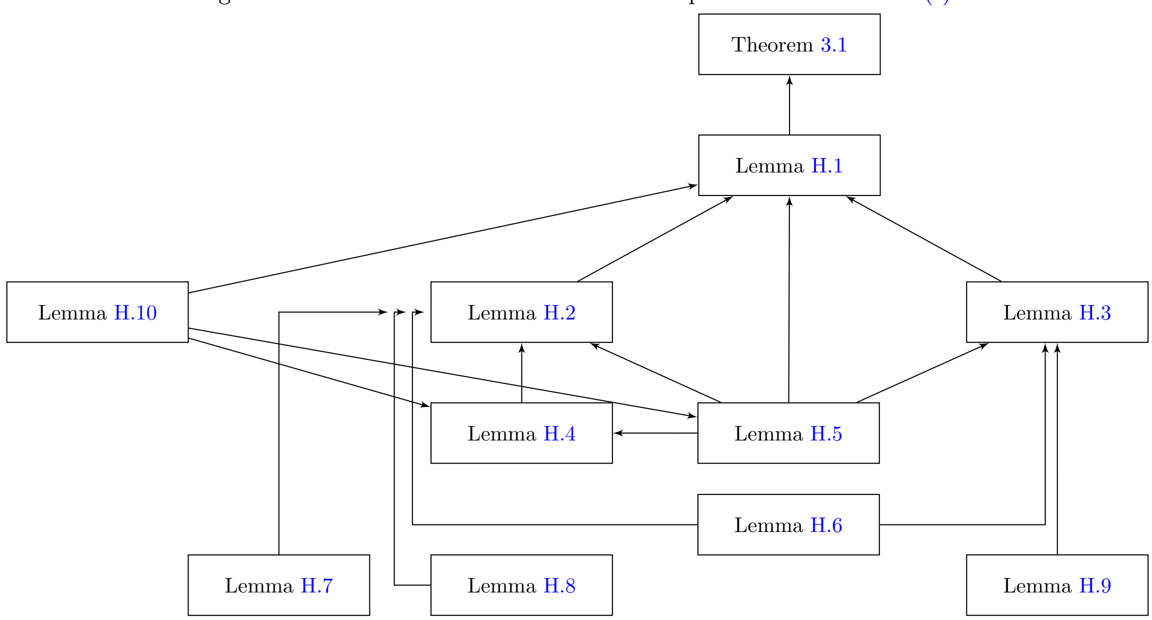
| | | | |
|---|---|---|---|
| $\mathbb{G}_{n,j}(\cdot)$ | $=$ | $\frac{\sqrt{n}(\bar{m}_{n,j}(\cdot) - E_P(m_j(X_i,\cdot)))}{\sigma_{P,j}(\cdot)},\ j = 1, \dots, J$ | Sample empirical process. |
| $\mathbb{G}^b_{n,j}(\cdot)$ | $=$ | $\frac{\sqrt{n}(\bar{m}^b_{n,j}(\cdot) - \bar{m}_{n,j}(\cdot))}{\hat{\sigma}_{n,j}(\cdot)},\ j = 1, \dots, J$ | 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. $\theta$, 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 H.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. |
| $\varphi^*_j(\xi)$ | $=$ | $\begin{cases} \varphi_j(\xi) & \pi_{1,j} = 0 \\ -\infty & \pi_{1,j} < 0 \\ 0 & j = J_1 + 1, \dots, J. \end{cases}$ | Infeasible GMS that is less conservative than $\varphi_j$. |
| $u_{n,j,\theta_n}(\lambda)$ | $=$ | $\{\mathbb{G}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\}(1 + \eta_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}))$ | Mean value expansion of nonlinear constraints with sample empirical process and "oracle" GMS, with $\bar{\theta}_n$ componentwise between $\theta_n$ 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 $p'\lambda = 0$. |
| $c_{\pi^*}$ | $=$ | $\inf\{c \in \mathbb{R}_+ : \Pr(\mathfrak{W}(c) \neq \emptyset) \geq 1 - \alpha\}.$ | 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 $\theta$. |
| $\hat{\sigma}^M_{n,j}(\theta)$ | $=$ | $\hat{\mu}_{n,j}(\theta)\hat{\sigma}_{n,j}(\theta) + (1 - \hat{\mu}_{n,j}(\theta))\hat{\sigma}_{n,j+R_1}(\theta)$ | Weighted sum of the estimators of the standard deviations of paired inequalities |

TABLE G.2

HEURISTICS FOR THE ROLE OF EACH LEMMA IN THE PROOF OF THEOREM 3.1. NOTES: (I) UNIFORMITY IN THEOREM 3.1 IS ENFORCED ARGUING ALONG SUBSEQUENCES; (II) WHEN NEEDED, RANDOM VARIABLES ARE REALIZED ON THE SAME PROBABILITY SPACE AS SHOWN IN LEMMA H.1 AND LEMMA H.17 (SEE APPENDIX H.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 (G.3)-(G.4) BELOW; (IV) ALL RESULTS HOLD FOR ANY $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$.

| | |
|---|---|
| Theorem 3.1 | $P_n(p'\theta_n \in CI) \geq P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}(\theta_n)) \neq \emptyset\right)$.<br>Coverage is conservatively estimated by the probability that $U_n$ is nonempty. |
| Lemma H.1 | $\liminf P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}(\theta_n)) \neq \emptyset\right) \geq 1 - \alpha$. |
| Lemma H.2 | $P_n(U(\theta_n, c_n^I(\theta_n)) \neq \emptyset, \mathfrak{W}(c_{\pi^*}) = \emptyset) + P_n(U(\theta_n, c_n^I(\theta_n)) = \emptyset, \mathfrak{W}(c_{\pi^*}) \neq \emptyset) = o_{\mathcal{P}}(1)$.<br>Argued by comparing $U_n$ and its limit $\mathfrak{W}$ (after coupling). |
| Lemma H.3 | $P_n^*(V_n^I(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset) \to 0$ and $c_n^I(\theta'_n) \xrightarrow{P_n} c_{\pi^*}$ if $c_{\pi^*} > 0$.<br>The bootstrap critical value that uses the less conservative GMS yileds a convergent critical value. |
| Lemma H.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 $v^I_{n,j,\theta'_n}$.<br>The criterion functions entering $U_n$ and $\mathfrak{W}$ converge to each other. |
| Lemma H.5 | Local-to-binding constraints are selected by GMS uniformly over the $\rho$-box (intuition: $\rho n^{-1/2} = o_{\mathcal{P}}(\kappa_n^{-1})$),<br>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 H.6 | $\forall \eta > 0\ \exists \delta > 0, : \Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}^{-\delta}(c) = \emptyset\}) < \eta$, and similarly for $V_n^I$.<br>It is unlikely that these sets are nonempty but become empty upon slightly tightening stochastic constraints. |
| Lemma H.7 | Intersections of constraints whose gradients are almost linearly dependent are unlikely to realize inside $\mathfrak{W}$.<br>Hence, we can ignore irregularities that occur as linear dependence is approached. |
| Lemma H.8 | If there are weakly more equality constraints than parameters, then $c$ is uniformly bounded away from zero.<br>This simplifies some arguments. |
| Lemma H.9 | If two paired inequalities are local to binding, then they are also asymptotically identical up to sign.<br>This justifies "merging" them. |
| Lemma H.10 | $\eta_{n,j}(\cdot)$ converges to zero uniformly in $P$ and $\theta$. |

Figure G.1: Structure of Lemmas used in the proof of Theorem 3.1-(I).



## G. *Proof of Theorem 3.1*

### G. *Main Proofs*

**Proof of Theorem 3.1-(I).**

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 $\gamma_{1,P} = (\gamma_{1,P,1}, \ldots, \gamma_{1,P,J})$ with

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

$\gamma_{2,P} = (s(p, \Theta_I(P)), vech(\Omega_P(\theta)), vec(D_P(\theta)))$, 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

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

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

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

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

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

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

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

The projection of $\theta_n$ 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 & \left\{ \begin{matrix} \inf p'\vartheta \\ \text{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{matrix} \right\} \le p'\theta_n \le \left\{ \begin{matrix} \sup p'\vartheta \\ \text{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{matrix} \right\} \\
\Leftrightarrow & \left\{ \begin{matrix} \inf_\lambda p'\lambda \\ \text{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{matrix} \right\} \le 0 \\
\le & \left\{ \begin{matrix} \sup_\lambda p'\lambda \\ \text{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{matrix} \right\} \qquad \text{(G.6)} \\
\Leftrightarrow & \left\{ \begin{matrix} \inf_\lambda p'\lambda \\ \text{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{matrix} \right\} \le 0 \\
\le & \left\{ \begin{matrix} \sup_\lambda p'\lambda \\ \text{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{matrix} \right\}, \qquad \text{(G.7)}
\end{aligned}$$

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

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

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

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

**Step 2.** This step is used only when Assumption E.3-2 is invoked. When this assumption is invoked, recall that in equation (2.6) we use the estimator specified in Lemma H.10 equation (H.196) for $\sigma_{P,j}, j = 1, \ldots, 2R_1$ (with $R_1 \le J_1/2$ defined in the statement of the assumption). In equation (2.11) we use the sample analog estimators of $\sigma_{P,j}$ for all $j = 1, \ldots, J$. To keep notation manageable, we explicitly denote the estimator used in (2.6) 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 $j = 1, \ldots, R_1$ such that

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

where $\pi^*_1$ is defined in (G.5). Let $E_j$ be the statement $\gamma_{1,P_n,j}(\theta_n) = \gamma_{1,P_n,j+R_1}(\theta_n) = 0$ and let

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

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

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

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

with the following weighted sum of the paired inequalities

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

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

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

with

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

It then follows from Assumption E.3-2 that these replacements are conservative because

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

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

**Step 3.** Next, we make the following comparisons:

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

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

For any constraint $j$ for which $\pi^*_{1,j} = 0$, (G.17) yields that replacing $\sqrt{n}\gamma_{1,P_n,j}(\theta_n)$ in (G.9) with $\pi^*_{1,j}$ introduces a conservative distortion. Under Assumption E.3-2, for any $j$ such that (G.10) holds, the substitutions in (G.14) and (G.16) yield

$$\tilde{\mu}_j\sqrt{n}\gamma_{1,P_n,j}(\theta_n)(1+\eta_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}})) - \tilde{\mu}_{j+R_1}\sqrt{n}\gamma_{1,P_n,j+R_1}(\theta_n)(1+\eta_{n,j+R_1}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}})) = 0, \tag{G.19}$$

and therefore replacing this term with $\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1}$ is inconsequential.

For any $j$ for which $\pi^*_{1,j} = -\infty$, (G.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 E.4 (i)-(ii), and $\lambda \in B^d_{n,\rho}$, it follows that

$$\rho D_{P_n,j}(\bar{\theta}_n)\lambda \leq \rho\sqrt{d}(\|D_{P_n,j}(\bar{\theta}_n) - D_{P_n,j}(\theta_n)\| + \|D_{P_n,j}(\theta_n)\|) \leq \rho\sqrt{d}(\rho M/\sqrt{n} + \bar{M}), \tag{G.20}$$

where $\bar{M}$ and $M$ are as defined in Assumption E.4-(i) and (ii) respectively, and we used that $\bar{\theta}_n$ lies component-wise between $\theta_n$ and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$. Using that $\mathbb{G}_{n,j}$ is asymptotically tight by Assumption E.5, we have that for any $\tau > 0$, there exists a $T > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$,

$$P_n\Big(\max_{j:\pi^*_{1,j}=-\infty}\{\mathbb{G}_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}})) \leq 0, \quad \forall\lambda \in B^d_{n,\rho}\Big) > 1 - \tau/2. \tag{G.21}$$

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 $\left|\frac{\sqrt{n}}{\kappa_n}\gamma_{1P_nj}(\theta_n) - \pi_{1j}\right| \leq \epsilon$, for all

$n \geq N_2$. Choose $\epsilon > 0$ such that $\pi_{1j} + \epsilon < 0$. Let $N = \max\{N_1, N_2\}$. Then we have

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

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

$$\begin{aligned}
&P_n\Bigg(\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\leq 0\\
&\qquad\leq\left\{\begin{array}{c}\sup_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\sqrt{n}\gamma_{1,P_n,j}(\theta_n)\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\Bigg) \qquad \text{(G.22)}\\
&\geq P_n\Bigg(\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\leq 0\\
&\qquad\leq\left\{\begin{array}{c}\sup_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\Bigg)-\tau/2. \qquad \text{(G.23)}
\end{aligned}$$

Since the choice of $\tau$ is arbitrary, the limit of the term in (G.22) is not smaller than the limit of the first term in (G.23). Hence, we continue arguing for the event whose probability is evaluated in (G.23).

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 (G.23) is implied by the event

$$\begin{aligned}
&\left\{\begin{array}{c}\inf_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\inf_{\lambda\in B^d_{n,\rho}}\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\}\leq 0\\
&\qquad\leq\left\{\begin{array}{c}\sup_\lambda p'\lambda\\ \text{s.t. } \lambda\in B^d_{n,\rho},\\ \{\mathbb{G}_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\pi^*_{1,j}\}(1+\eta_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}}))\leq\inf_{\lambda\in B^d_{n,\rho}}\hat c_n(\theta_n+\frac{\lambda\rho}{\sqrt{n}}),\forall j\end{array}\right\} \qquad \text{(G.24)}
\end{aligned}$$

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

$$u_{n,j,\theta_n}(\lambda) \equiv \left\{\mathbb{G}_{n,j}(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})+\rho D_{P_n,j}(\bar\theta_n)\lambda+\pi^*_{1,j}\right\}\left(1+\eta_{n,j}(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})\right), \qquad \text{(G.25)}$$

where under Assumption E.3-2 when $\pi^*_{1,j} = 0$ and $\pi^*_{1,j+R_1} = 0$ the substitutions of equation (G.13) with equation (G.14) and of equation (G.15) with equation (G.16) have been performed. Let

$$U_n(\theta_n, c) \equiv \left\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u_{n,j,\theta_n}(\lambda) \le c,\ \forall j = 1, \dots, J\right\}, \tag{G.26}$$

and define

$$\hat{c}_{n,\rho} \equiv \inf_{\lambda \in B^d_{n,\rho}} \hat{c}_n(\theta + \tfrac{\lambda\rho}{\sqrt{n}}). \tag{G.27}$$

Then by (G.24) and the definition of $U_n$, we obtain

$$P_n(p'\theta_n \in CI_n) \ge P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}) \ne \emptyset\right). \tag{G.28}$$

By passing to a further subsequence, we may assume that

$$D_{P_n}(\theta_n) \to D, \tag{G.29}$$

for some $J \times d$ matrix $D$ such that $\|D\| \le M$ and $\Omega_{P_n} \xrightarrow{u} \Omega$ for some correlation matrix $\Omega$. By Lemma 2 in Andrews and Guggenberger (2009) and Assumption E.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 $\Omega$. By Lemma H.1,

$$\liminf_{n\to\infty} P_n\left(U_n(\theta_n, \hat{c}_{n,\rho}) \ne \emptyset\right) \ge 1 - \alpha. \tag{G.30}$$

The conclusion of the theorem then follows from (G.2), (G.3), (G.28), and (G.30). *Q.E.D.*

**Proof of Theorem 3.1-(II).**

The result follows immediately from the same steps as in the proof of Theorem 3.1-(I). *Q.E.D.*

**Proof of Theorem 3.1-(III)**

The argument of proof is the same as for Theorem 3.1-(I), with the following modification. Take $(P_n, \theta_n)$ as defined following equation (G.4). Then $f(\theta_n)$ is covered when

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

where we took a mean value expansion yielding

$$f(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) = f(\theta_n) + \frac{\rho}{\sqrt{n}}\nabla f(\tilde{\theta}_n)\lambda, \tag{G.31}$$

for $\tilde{\theta}_n$ a mean value that lies componentwise between $\theta_n$ and $\theta_n + \frac{\lambda\rho}{\sqrt{n}}$, and we used that the sign of the last term in (G.31) is the same as the sign of $\nabla f(\tilde{\theta}_n)\lambda$. With the objective function in (G.31) so redefined, all expression in the proof of Theorem 3.1-(I) up to (G.25) continue to be valid. We can then redefine the set $U_n(\theta_n, c)$ in (G.26) as

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

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 $V^I_n(\theta'_n, \hat{c}^f_n(\theta'_n))$, 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 (H.17) is unchanged. This yields the result, noting that by the assumption $\|\nabla f(\tilde{\theta}_n) - \nabla f(\theta'_n)\| = O_{\mathcal{P}}(\rho/\sqrt{n})$. *Q.E.D.*

### G. *Proof of Theorem 3.1-(I) with High Level Assumption E.6 Replacing Assumption E.3, and Dropping the ρ-Box Constraints Under Assumption E.7*

LEMMA G.1 *Suppose that Assumption E.1, E.2, E.4 and E.5 hold.*

*(I) Let also Assumption E.6 hold. Let $0 < \alpha < 1/2$. Then,*

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

*(II) Let also Assumption E.7 and either Assumption E.3 or E.6 hold. Let $\hat{c}_n = \inf\{c \in \mathbb{R}_+ : P^*(\{\Lambda^b_n(\theta, +\infty, c) \cap \{p'\lambda = 0\}\} \neq \emptyset) \geq 1 - \alpha\}$, where $\Lambda^b_n$ is defined in equation (2.11) 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 (2.6). Then*

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

PROOF: We establish each part of the Lemma separately.

**Part I.** This part of the lemma replaces Assumptions E.3 with Assumption E.6. Hence we establish the result by showing that all claims that were made under Assumption E.3 remain valid under Assumption E.6. We proceed in steps.

Step 1. Revisiting the proof of Lemma H.6, equation (H.139).

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

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

$$\mathfrak{W}(c - \delta) \equiv \left\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \leq c - \delta,\ \forall j = 1, \ldots, J\right\}, \tag{G.32}$$

where we emphasize that the set $\mathfrak{W}(c-\delta)$ is obtained by a $\delta$-contraction of all constraints, including those indexed by $j = J_1 + 1, \ldots, J$. By Assumption E.6, for any $\eta > 0$ there exists a $\delta$ 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 H.2.

Case 1 of Lemma H.2 is unaltered. Case 2 of Lemma H.2 follows from the same argument as used in Case 1 of Lemma H.2, because under Assumption E.6 as shown in step 1 of this proof all inequalities are tightened. In Case 3 of Lemma H.2 the result in (G.30) holds automatically by Assumption E.6-(ii). (As a remark, Lemmas H.7-H.8 are no longer needed to establish Lemma H.2.)

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

In summary, the desired result follows by applying Lemma H.1 in the proof of Theorem 3.1-(I) as Lemmas H.2, H.3 and H.6 remain valid, Lemmas H.4, H.5, H.10 and the Lemmas in Appendix H.3 are unaffected, and Lemmas H.7, H.8, H.9 are no longer needed.

**Part II.** This is established by adapting the proof of Theorem 3.1-(I) 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 $\rho$. 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 (G.9). Indeed, (G.9) is replaced by:

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

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

yielding a new definition of the set $U_n^*$ as

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

Subsequent uses of $\rho$ in the main proof use that $\|\lambda\| \leq \sqrt{d}\rho = O_{\mathcal{P}}(1)$. For example, consider the argument following equation (G.20) or the argument just preceding equation (G.30), and so on. All these continue to go through because $\bar{\mathfrak{W}}^*(\bar{c}) = O(1)$ by assumption.

Similar uses occur in Lemma H.1. The next major adaptation is that in (H.27) and (H.28): we again drop $\rho$ 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 H.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 H.7. This requires straightforward adaptation of algebra as $\bar{\mathfrak{W}}(\bar{c})$ is only stochastically and not deterministically bounded.

Finally, in Lemma H.3 we remove the $\rho$-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 $\lambda$ is uniformly stochastically bounded in expressions like (H.98), we verify that the modified $\tilde{V}_n^b$ and $\tilde{V}_n^I$ inherit the property in Assumption E.7. 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 E.7 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 H.7 and the fact that $\hat{D}_n(\theta_n') \xrightarrow{P} D$ by Assumption E.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 H.3 goes through. *Q.E.D.*

## G. *An Extension of Theorem 3.1*

In this subsection, we establish that, under the assumptions of Theorem 3.1, we actually have

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(p'\theta \in \{p'\vartheta : \vartheta \in \mathcal{C}_n(\hat{c}_n)\}) \geq 1 - \alpha. \tag{G.33}$$

In words, the mathematical projection of $\mathcal{C}_n(\hat{c}_n)$, which will asymptotically pick up gaps in the projection of $\Theta_I$, is a uniformly asymptotically valid confidence region. This strengthens Theorem 3.1 because $\{p'\vartheta : \vartheta \in \mathcal{C}_n(\hat{c}_n)\} \subseteq CI_n$.

To prove this extension, we modify the proof of Theorem 3.1 after (G.5) as follows: The projection of $\theta_n$ is covered when

$$\exists \vartheta \in \Theta : p'\vartheta = p'\theta_n, \tfrac{\sqrt{n}\bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \leq \hat{c}_n(\vartheta), \forall j \tag{G.34}$$

$$\iff \exists \lambda \in \tfrac{\sqrt{n}}{\rho}(\Theta - \theta_n) : p'\lambda = 0, \frac{\sqrt{n}\bar{m}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})}{\hat{\sigma}_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}})} \leq \hat{c}_n(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}), \forall j \tag{G.35}$$

$$\iff \exists \lambda \in \tfrac{\sqrt{n}}{\rho}(\Theta - \theta_n) : \tag{G.36}$$

$$p'\lambda = 0, \big(\mathbb{G}_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \sqrt{n}\gamma_{1,P_n,j}(\theta_n)\big)\big(1 + \eta_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}})\big) \leq \hat{c}_n(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}), \forall j \tag{G.37}$$

where the last line corresponds to (G.7) and intermediate steps that are exactly analogous to the previous proof were skipped. Subsequent proof steps go through as before until, comparing (G.26) to (G.37), we find (compare to (G.28), noting the change from inequality to equality)

$$P_n\big(p'\theta_n \in \{p'\vartheta : \vartheta \in \mathcal{C}_n(\hat{c}_n)\}\big) = P_n\big(U_n(\theta_n, \hat{c}_{n,\rho}) \neq \emptyset\big). \tag{G.38}$$

The proof then continues as before.

## H. Auxiliary Lemmas

### *H. Lemmas Used to Prove Theorem 3.1*

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 3.1-(I). That is, along $(P_n, \theta_n)$, one has

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

$$\Omega_{P_n} \xrightarrow{u} \Omega, \tag{H.2}$$

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

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

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

where $\pi^*_{1,j}$ is defined in (G.5) and we used Lemma H.5. Under Assumption E.3-2 if

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

we replace the constraints

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

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

with

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

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

where

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

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

When Assumption E.3-2 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 (2.11) as described on p.11 of the paper; when it is invoked with a GMS function $\varphi$ that is smooth in its argument, replace them, respectively, with

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

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

where

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

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

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

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

Due to the substitutions in equations (H.6)-(H.9), the paired inequalities (i.e., inequalities for which (H.5) holds under Assumption E.3-2) 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

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

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

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

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

LEMMA H.1 *Let Assumptions E.1, E.2, E.3, E.4, and E.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 $\kappa_n^{-1}\sqrt{n}\gamma_{1,P_n,j}(\theta_n) \to \pi_{1j} \in \mathbb{R}_{[-\infty]}$, $j = 1, \dots, J$, $\Omega_{P_n} \xrightarrow{u} \Omega$, and $D_{P_n}(\theta_n) \to D$. Then,*

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

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

$$\hat{c}_n(\theta_n') \ge c_n^I(\theta_n') + o_{P_n}(1), \tag{H.20}$$

and $c_n^I(\theta_n') \xrightarrow{P_n} c_{\pi^*}$ for any $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$. The construction is as follows. When $c_{\pi^*} = 0$, let $c_n^I(\theta_n') = 0$ for all $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$, and hence $c_n^I(\theta_n') \xrightarrow{P_n} c_{\pi^*}$. If $c_{\pi^*} > 0$, let $c_n^I(\theta_n) \equiv$

$\inf\{c \in \mathbb{R}_+ : P_n^*(V_n^I(\theta_n, c)) \geq 1-\alpha\}$, where $V_n^I$ is defined as in Lemma H.3. By Lemma H.3 (iii), this critical value sequence satisfies (H.20) with probability approaching 1. Further, by Lemma H.3 (ii), $c_n^I(\theta_n') \xrightarrow{P_n} c_{\pi^*}$ for any $\theta_n' \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$.

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

$$c_{n,\rho}^I(\theta) \equiv \inf_{\lambda \in B_{n,\rho}^d} c_n^I(\theta + \tfrac{\lambda\rho}{\sqrt{n}}). \tag{H.21}$$

Since the $o_{P_n}(1)$ term in (H.20) does not affect the argument below, we redefine $c_{n,\rho}^I(\theta_n)$ as $c_{n,\rho}^I(\theta_n) + o_{P_n}(1)$. By (H.20) and simple addition and subtraction,

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

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

$$\big(\mathbb{G}_n(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}), \eta_n(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}), c_{n,\rho}^I(\theta_n)\big) \xrightarrow{d} (\mathbb{Z}, 0, c_{\pi^*}). \tag{H.23}$$

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 $(\mathbb{G}_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), \eta_n(\theta_n + \frac{\lambda\rho}{\sqrt{n}}), c_{n,\rho}^I(\theta_n))$ 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 (G.25) and (H.4) as

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

$$\mathfrak{w}_j^*(\lambda) \equiv \mathbb{Z}_j^* + \rho D_j\lambda + \pi_{1,j}^*. \tag{H.25}$$

where we used that by Lemma H.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 (H.6)-(H.9) for the case that $\pi_{1,j}^* = 0 = \pi_{1,j+R_1}^*$. Similarly, we define analogs to the sets in (G.26) and (H.16) as

$$U_n^*(\theta_n, c_n^*) \equiv \big\{\lambda \in B_{n,\rho}^d : p'\lambda = 0 \cap u_{n,j,\theta_n}^*(\lambda) \leq c_n^*,\ \forall j = 1, \dots, J\big\}, \tag{H.26}$$

$$\mathfrak{W}^*(c_{\pi^*}) \equiv \big\{\lambda \in \mathfrak{B}_\rho^d : p'\lambda = 0 \cap \mathfrak{w}_j^*(\lambda) \leq c_{\pi^*},\ \forall j = 1, \dots, J\big\}. \tag{H.27}$$

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

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

By the definition of $c_{\pi^*}$, we have $\mathbf{P}(\mathfrak{W}^*(c_{\pi^*}) \neq \emptyset) \geq 1-\alpha$. Therefore, we are left to show that the second term on the right hand side of (H.28) tends to 0 as $n \to \infty$.

Define

$$\mathcal{J}^* \equiv \{j = 1, \dots, J : \pi_{1,j}^* = 0\}. \tag{H.29}$$

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

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

with probability 1, and hence

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

This in turn implies that

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

where we used $|\mathbf{P}(A) - \mathbf{P}(B)| \leq \mathbf{P}(A\Delta B) \leq 1 - \mathbf{P}(A \cap B)$ for any pair of events $A$ and $B$. Hence, the term in the square brackets in (H.28) is 0.
**Case 2.** Now consider the case that $\mathcal{J}^* \neq \emptyset$. We show that the term in the square brackets in (H.28) converges to 0. To that end, note that for any events $A, B$,

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

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 (H.33) converges to zero. But this is guaranteed by Lemma H.2. Therefore, the conclusion of the lemma follows. *Q.E.D.*

Lemma H.2 *Let Assumptions E.1, E.2, E.3, E.4, and E.5 hold. Let $(P_n, \theta_n)$ have the almost sure representations given in Lemma H.1, and let $\mathcal{J}^*$ be defined as in (H.29). Assume that $\mathcal{J}^* \neq \emptyset$. Then for any $\eta > 0$, there exists $N \in \mathbb{N}$ such that*

$$\mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big) \leq \eta/2, \tag{H.34}$$

$$\mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) = \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) \neq \emptyset\}\Big) \leq \eta/2, \tag{H.35}$$

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

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

$$\Big(\mathbb{G}^*_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) + \rho D_{P_n,j}(\bar{\theta}_n)\lambda + \pi^*_{1,j}\Big)(1 + \eta^*_{n,j}) \leq c^*_n,$$
$$\mathbb{Z}^*_j + \rho D_j \lambda + \pi^*_{1,j} \leq c_{\pi^*}$$

are satisfied with probability approaching one by similar arguments as in (G.21). Hence, we can redefine the sets of interest as

$$U^*_n(\theta_n, c^*_n) \equiv \big\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u^*_{n,j,\theta_n}(\lambda) \leq c^*_n,\ \forall j \in \mathcal{J}^*\big\}, \tag{H.36}$$

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

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

$$A_n \equiv \left\{ \sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} \left| (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - (\mathfrak{w}^*_j(\lambda) - c_{\pi^*}) \right| \geq \delta \right\}. \tag{H.38}$$

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

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

Define the sets of $\lambda$s, $U^{*,+\delta}_n$ and $\mathfrak{W}^{*,+\delta}$ by relaxing the constraints shaping $U^*_n$ and $\mathfrak{W}^*$ by $\delta$:

$$U^{*,+\delta}_n(\theta_n, c) \equiv \{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u^*_{n,j,\theta_n}(\lambda) \leq c + \delta, \ j \in \mathcal{J}^*\}, \tag{H.40}$$

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

Compared to the set in equation (H.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 $\{1, \ldots, J_1\}$. Next, define the event $L_n \equiv \{U^*_n(\theta_n, c^*_n) \subset \mathfrak{W}^{*,+\delta}(c_{\pi^*})\}$ and note that $A^c_n \subseteq L_n$.

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

$$\begin{aligned} \mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big) &\leq \mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) \neq \emptyset\} \cap \{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) = \emptyset\}\Big) \\ &\quad + \mathbf{P}\Big(\{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) \neq \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) = \emptyset\}\Big), \end{aligned} \tag{H.42}$$

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 (H.42) can further be bounded as

$$\begin{aligned} \mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) \neq \emptyset\} \cap \{\mathfrak{W}^{*,+\delta}(c_{\pi^*}) = \emptyset\}\Big) &\leq \mathbf{P}\Big(\{U^*_n(\theta_n, c^*_n) \not\subseteq \mathfrak{W}^{*,+\delta}(c_{\pi^*})\}\Big) \\ &= \mathbf{P}(L^c_n) \leq \mathbf{P}(A_n) < \eta/2, \ \forall n \geq N \ , \end{aligned} \tag{H.43}$$

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

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

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

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

**Case 1.** Suppose first that $J_2 = 0$ (recalling that under Assumption E.3-2 this means that there is no $j = 1, \ldots, R_1$ such that $\pi^*_{1,j} = 0 = \pi^*_{1,j+R_1}$), and hence one has only moment inequalities. In this case, by (H.36) and (H.37), one may write

$$U^*_n(\theta_n, c) \equiv \big\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap u^*_{n,j,\theta_n}(\lambda) \leq c, \ j \in \mathcal{J}^*\big\}, \tag{H.45}$$

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

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

$$R_{2n} \equiv \{\mathfrak{W}^{*,-\delta}(c_{\pi^*}) \subset U^*_n(\theta_n, c^*_n)\}, \tag{H.47}$$

and note that $A_n^c \subseteq R_{2n}$. The result in equation (H.35) then follows by Lemma H.6 using again similar steps to (H.42)-(H.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 $\delta$ the inequality constraints as well as each of the two opposing inequalities obtained from the equality constraints. That is,

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

that is, the same set as in (H.139) with $\mathfrak{w}_j^*(\lambda)$ replacing $\mathfrak{w}_j(\lambda)$ and defining the set using only inequalities in $\mathcal{J}^*$. Note that, by Lemma H.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 $c_{\pi^*}$ 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 H.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 (H.47) now defined using the tightened constraint set $\mathfrak{W}^{*,-\delta}(c_{\pi^*})$ being defined as in (H.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 $\mathbf{P}$),

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

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

$$\mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) = \emptyset\} \cap \{\mathfrak{W}^*(c_{\pi^*}) \neq \emptyset\}\Big) \tag{H.50}$$

$$\leq \mathbf{P}\Big(\{U_n^*(\theta_n, c_n^*) = \emptyset\} \cap \{\mathfrak{W}^{*,-\delta}(0) \neq \emptyset\}\Big) + \mathbf{P}\Big(\{\mathfrak{W}^{*,-\delta}(0) = \emptyset\} \cap \{\mathfrak{W}^*(0) \neq \emptyset\}\Big), \tag{H.51}$$

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

$$\mathbf{P}\Big(\{\mathfrak{W}^{*,-\delta}(0) = \emptyset\} \cap \{\mathfrak{W}^*(0) \neq \emptyset\}\Big) < \eta/3 \text{ for all } n \geq N. \tag{H.52}$$

Therefore, the second term on the right hand side of (H.51) can be made arbitrarily small.

We now consider the first term on the right hand side of (H.51). Let $g$ be a $J + 2d + 2$ vector with

$$g_j = \begin{cases} -\mathbb{Z}_j, & j \in \mathcal{J}^*, \\ 0, & j \in \{1, \ldots, J\} \setminus \mathcal{J}^*, \\ 1, & j = J+1, \ldots, J+2d, \\ 0, & j = J+2d+1, J+2d+2, \end{cases} \tag{H.53}$$

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 (H.36)-(H.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

$$K \equiv \begin{bmatrix} [\rho D_j]_{j=1}^{J_1+J_2} \\ [-\rho D_{j-J_2}]_{j=J_1+J_2+1}^{J} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}. \tag{H.54}$$

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 $\mathfrak{B}^d_\rho$ otherwise, see, e.g., Molchanov (2005, Definition 1.2.2).

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

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \eta, \tag{H.55}$$

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

$$\mathbf{P}\Big(\big|\det\left(K^{\mathcal{C}}\right)^{-1}\big| \leq M\Big) \geq 1-\eta, \ \forall n \geq N. \tag{H.56}$$

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

$$g_{n,j}(\theta + \lambda/\sqrt{n}) \equiv \begin{cases} c_n^*/(1+\eta_{n,j}^*) - \mathbb{G}_{n,j}^*(\theta + \frac{\lambda\rho}{\sqrt{n}}) & \text{if } j \in \mathcal{J}^*, \\ 0, & \text{if } j \in \{1, \ldots, J\} \setminus \mathcal{J}^*, \\ 1, & \text{if } j = J+1, \ldots, J+2d, \\ 0, & \text{if } j = J+2d+1, J+2d+2, \end{cases} \tag{H.57}$$

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

$$K_P(\theta, \rho) \equiv \begin{bmatrix} [\rho D_{P,j}(\theta)]_{j=1}^{J_1+J_2} \\ [-\rho D_{P,j-J_2}(\theta)]_{j=J_1+J_2+1}^{J} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}. \tag{H.58}$$

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

$$\phi_n(\lambda) \equiv \left(K_{P_n}^{\mathcal{C}}(\bar{\theta}(\theta_n, \lambda), \rho)\right)^{-1} g_n^{\mathcal{C}}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}), \tag{H.59}$$

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

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

$$\begin{aligned}\|\phi_n(\lambda)-\phi_n(\lambda')\| &= \Big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})-\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda'),\rho)\big)^{-1}g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|\\ &\le \Big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}\Big\|_2\Big\|g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})-g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|\\ &\quad+\Big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}-\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda'),\rho)\big)^{-1}\Big\|_2\Big\|g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|,\end{aligned}\tag{H.60}$$

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

By Assumption E.5 (ii), for any $\eta>0$, $k>0$, there is $N\in\mathbb{N}$ such that

$$\mathbf{P}\left(\Big\|g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})-g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|\le k\|\lambda-\lambda'\|\right)\tag{H.61}$$

$$=\mathbf{P}\left(\Big\|\mathbb{G}^{*,\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})-\mathbb{G}^{*,\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|\le k\|\lambda-\lambda'\|\right)\ge 1-\eta,\ \forall n\ge N.\tag{H.62}$$

Moreover, by arguing as in equation (G.21), for any $\eta$ there exist $0<L<\infty$ and $N\in\mathbb{N}$ such that $\forall n\ge N$

$$\mathbf{P}\left(\sup_{\lambda'\in B^d}\Big\|g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda'\rho}{\sqrt{n}})\Big\|\le L\right)\ge 1-\eta.\tag{H.63}$$

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

$$\mathbf{P}\big(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2\le L\big)\ge 1-\eta,\ \forall n\ge N,\tag{H.64}$$

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

$$\|K^{-1}-\tilde K^{-1}\|_2\le\frac{\|\tilde K^{-1}(K-\tilde K)\|_2}{1-\|\tilde K^{-1}(K-\tilde K)\|_2}\|\tilde K^{-1}\|_2.\tag{H.65}$$

By the assumption that $D_{P_n}(\theta_n)\to D$ and Assumption E.4, for any $\eta>0$, there exists $N\in\mathbb{N}$ such that

$$\sup_{\lambda\in B^d}\|K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)-K^{\mathcal{C}}\|_2\le\eta,\ \forall n\ge N.\tag{H.66}$$

By (H.65), 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

$$\begin{aligned}&\mathbf{P}\big(\sup_{\lambda\in B^d}\big\|(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho))^{-1}\big\|_2\le 2L_1\big)\\ &\ge\mathbf{P}\big(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2+\sup_{\lambda\in B^d}\|K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)^{-1}-(K^{\mathcal{C}})^{-1}\|_2\le 2L_1\big)\\ &\ge\mathbf{P}\left(\big\|\big(K^{\mathcal{C}}\big)^{-1}\big\|_2\le L_1,\frac{\big\|(K^{\mathcal{C}})^{-1}\big\|_2^2}{1-\big\|(K^{\mathcal{C}})^{-1}(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)-K^{\mathcal{C}})\big\|_2}\le L_2,\ \sup_{\lambda\in B^d}\|K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)-K^{\mathcal{C}}\|_2\le\frac{L_1}{L_2}\right)\\ &\ge 1-2\eta,\ \forall n\ge N,\end{aligned}\tag{H.67}$$

Again by applying (H.65), for any $k>0$, there exists $N\in\mathbb{N}$ such that

$$\mathbf{P}\Big(\big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda))\big)^{-1}-\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda'))\big)^{-1}\big\|_2\le k\|\lambda-\lambda'\|\Big) \tag{H.68}$$

$$\ge \mathbf{P}\Big(\sup_{\lambda\in B^d}\big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda))\big)^{-1}\big\|_2^2 M\rho\|\bar\theta(\theta_n,\lambda)-\bar\theta(\theta_n,\lambda')\|\le k\|\lambda-\lambda'\|\Big)\ge 1-\eta,\ \forall n\ge N, \tag{H.69}$$

where the first inequality follows from $\|K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda))-K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda'))\|_2\le M\rho\|\bar\theta(\theta_n,\lambda)-\bar\theta(\theta_n,\lambda')\|\le M\rho^2/\sqrt{n}\|\lambda-\lambda'\|$ by Assumption E.4 (ii), and the last inequality follows from (H.67).

By (H.60)-(H.63) and (H.67)-(H.69), it then follows that there exists $\beta\in[0,1)$ such that for any $\eta>0$, there exists $N\in\mathbb{N}$ such that

$$\mathbf{P}\left(|\phi_n(\lambda)-\phi_n(\lambda')|\le\beta\|\lambda-\lambda'\|,\quad \forall\lambda,\lambda'\in B^d\right)\ge 1-\eta,\ \forall n\ge N. \tag{H.70}$$

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

$$\mathbf{P}\left(\exists\lambda_n^f:\lambda_n^f=\phi_n(\lambda_n^f)\right)\ge 1-\eta,\ \forall n\ge N. \tag{H.71}$$

Next, define the mapping

$$\psi_n(\lambda)\equiv\left(K^{\mathcal{C}}\right)^{-1}g^{\mathcal{C}}. \tag{H.72}$$

This map is constant in $\lambda$ 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 (H.59) and (H.72) arguing as in (H.60), it follows that for any $\lambda\in B^d$,

$$\begin{aligned}\|\psi_n(\lambda)-\phi_n(\lambda)\|\le{}&\Big\|\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}\Big\|_2\Big\|g^{\mathcal{C}}-g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})\Big\|\\&+\Big\|\big(K^{\mathcal{C}}\big)^{-1}-\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}\Big\|_2\|g^{\mathcal{C}}\|.\end{aligned} \tag{H.73}$$

By (H.53) and (H.57)

$$\begin{aligned}\Big\|g^{\mathcal{C}}-g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})\Big\|&\le\max_{j\in\mathcal{J}^*}|-\mathbb{Z}^*_j-c^*_n/(1+\eta^*_{n,j})+\mathbb{G}^*_{n,j}(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})|\\&\le\max_{j\in\mathcal{J}^*}|\mathbb{Z}^*_j-\mathbb{G}^*_{n,j}(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})|+\max_{j\in\mathcal{J}^*}|c^*_n/(1+\eta^*_{n,j})|.\end{aligned} \tag{H.74}$$

We note that when Assumption E.3-2 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 $\mu_j$ were defined in (G.11)-(G.12) and (H.10)-(H.11) respectively. Moreover, $\mathbb{G}^*_{n,j}(\theta_n+\frac{\lambda\rho}{\sqrt{n}})\stackrel{a.s.}{\to}\mathbb{Z}^*$ and (H.49) implies $c^*_n\to 0$ so that we have

$$\sup_{\lambda\in B^d}\Big\|g^{\mathcal{C}}-g^{\mathcal{C}}_n(\theta_n+\tfrac{\lambda\rho}{\sqrt{n}})\Big\|\stackrel{a.s.}{\to}0. \tag{H.75}$$

Further, by (H.65), $D_{P_n}\to D$ and, Assumption E.4-(ii), for any $\eta>0$, there exists $N\in\mathbb{N}$ such that

$$\sup_{\lambda\in B^d}\Big\|\big(K^{\mathcal{C}}\big)^{-1}-\big(K^{\mathcal{C}}_{P_n}(\bar\theta(\theta_n,\lambda),\rho)\big)^{-1}\Big\|_2\le\eta,\ \forall n\ge N. \tag{H.76}$$

In sum, by (H.63), (H.67), and (H.74)-(H.76), for any $\eta,\nu>0$, there exists $N\geq\mathbb{N}$ such that

$$\mathbf{P}\left(\sup_{\lambda\in B^d}\|\psi_n(\lambda)-\phi_n(\lambda)\|<\nu\right)\geq 1-\eta,\ \forall n\geq\mathbb{N}. \tag{H.77}$$

Hence, for a specific choice of $\nu=\kappa(1-\beta)$, where $\beta$ is defined in equation (H.70), we have that $\sup_{\lambda\in B^d}\|\psi_n(\lambda)-\phi_n(\lambda)\|<\kappa(1-\beta)$ implies

$$\begin{aligned}\|\lambda_n^{\mathcal{C}}-\lambda_n^f\| &= \|\psi_n(\lambda_n^{\mathcal{C}})-\phi_n(\lambda_n^f)\|\\ &\leq \|\psi_n(\lambda_n^{\mathcal{C}})-\phi_n(\lambda_n^{\mathcal{C}})\|+\|\phi_n(\lambda_n^{\mathcal{C}})-\phi_n(\lambda_n^f)\|\\ &\leq \kappa(1-\beta)+\beta\|\lambda_n^{\mathcal{C}}-\lambda_n^f\| \end{aligned}\tag{H.78}$$

Rearranging terms, we obtain $\|\lambda_n^{\mathcal{C}}-\lambda_n^f\|\leq\kappa$. Note that by Assumptions E.4 (i) and E.5 (i), for any $\delta>0$, there exists $\kappa_\delta>0$ and $N\in\mathbb{N}$ such that

$$\mathbf{P}\Big(\sup_{\|\lambda-\lambda'\|\leq\kappa_\delta}|u^*_{n,j,\theta_n}(\lambda)-u^*_{n,j,\theta_n}(\lambda')|<\delta\Big)\geq 1-\eta,\ \forall n\geq\mathbb{N}. \tag{H.79}$$

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

$$\mathfrak{w}^*_j(\lambda_n^{\mathcal{C}})+\delta\leq 0,\ j\in\{1,\dots,J_1\}\cap\mathcal{J}^*. \tag{H.80}$$

Hence, by (H.39), (H.49), and (H.79)-(H.80), $\|\lambda_n^{\mathcal{C}}-\lambda_n^f\|\leq\kappa_{\delta/4}$, for each $j\in\{1,\dots,J_1\}\cap\mathcal{J}^*$ we have

$$u^*_{n,j,\theta_n}(\lambda_n^f)-c_n^*(\theta_n)\leq u^*_{n,j,\theta_n}(\lambda_n^{\mathcal{C}})-c_n^*(\theta_n)+\delta/4\leq\mathfrak{w}^*_j(\lambda_n^{\mathcal{C}})+\delta/2\leq 0. \tag{H.81}$$

For $j\in\{J_1+1,\dots,2J_2\}\cap\mathcal{J}^*$, 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 $n\geq N$ we have

$$\begin{aligned}\mathbf{P}\Big(\{U_n^*(\theta_n,c_n^*)=\emptyset\}\cap\{\mathfrak{W}^{*,-\delta}(0)\neq\emptyset\}\Big)&\leq\mathbf{P}\Big(\nexists\lambda_n^f\in U_n^*(\theta_n,c_n^*),\exists\lambda_n^{\mathcal{C}}\in\mathfrak{W}^{*,-\delta}(0)\Big)\\ &\leq\mathbf{P}\left(\left\{\sup_{\lambda\in B^d}\|\psi_n(\lambda)-\phi_n(\lambda)\|<\kappa_\delta(1-\beta)\cap A_n\right\}^c\right)\leq\eta/3,\end{aligned}\tag{H.82}$$

where $A^c$ denotes the complement of the set $A$, and the last inequality follows from (H.39) and (H.77). *Q.E.D.*

Lemma H.3 *Suppose Assumptions E.1, E.2, E.3, E.4, and E.5 hold. Let $\{P_n,\theta_n\}\in\{(P,\theta):P\in\mathcal{P},\theta\in\Theta_I(P)\}$ be a sequence satisfying (H.1)-(H.3). For each $j$, let*

$$v^I_{n,j,\theta_n}(\lambda)\equiv\mathbb{G}^b_{n,j}(\theta_n)+\rho\hat{D}_{n,j}(\theta_n)\lambda+\varphi^*_j(\hat{\xi}_{n,j}(\theta_n)), \tag{H.83}$$

$$\mathfrak{w}_j(\lambda)\equiv\mathbb{Z}_j+\rho D_j\lambda+\pi^*_{1,j}, \tag{H.84}$$

*where*

$$\varphi^*_j(\xi)=\begin{cases}\varphi_j(\xi) & \pi_{1,j}=0\\ -\infty & \pi_{1,j}<0\\ 0 & j=J_1+1,\dots,J.\end{cases} \tag{H.85}$$

*For each $c \geq 0$, define*

$$V_n^I(\theta_n, c) \equiv \{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap v^I_{n,j,\theta_n}(\lambda) \leq c, j = 1, \dots, J\}, \tag{H.86}$$

$$\mathfrak{W}(c) \equiv \big\{\lambda \in \mathfrak{B}^d_\rho : p'\lambda = 0 \cap \mathfrak{w}_j(\lambda) \leq c, \ \forall j = 1, \dots, J\big\}. \tag{H.87}$$

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

*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 $n$,*

$$P_n^*(V_n^I(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset) \to 0, \tag{H.88}$$

*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 $n$,*

$$\hat{c}_n(\theta'_n) \geq c_n^I(\theta'_n) + o_{P_n}(1). \tag{H.89}$$

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 H.15, in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_i\}_{i=1}^\infty$, and by Assumption E.4 $\|\hat{D}_n(\theta'_n) - D_{P_n}(\theta_n)\| \xrightarrow{p} 0$. Further, by Lemma H.5, $\hat{\xi}_{n,j}(\theta'_n) \xrightarrow{P_n} \pi_{1,j}$. Therefore,

$$(\mathbb{G}_n^b(\theta'_n), \hat{D}_n(\theta'_n), \hat{\xi}_n(\theta'_n)) | \{X_i\}_{i=1}^\infty \xrightarrow{d} (\mathbb{Z}, D, \pi_1). \tag{H.90}$$

for almost all sample paths $\{X_i\}_{i=1}^\infty$. By Lemma H.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 $(\tilde{\mathbb{G}}_n^b(\theta'_n), \tilde{D}_n, \tilde{\xi}_{n,j}) \xrightarrow{a.s.} (\tilde{\mathbb{Z}}, D, \pi_1)$ 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 $x^\infty$ for notational simplicity (see Appendix H.3 for details). Using this representation, define

$$\tilde{v}^I_{n,j,\theta'_n}(\lambda) \equiv \tilde{\mathbb{G}}^b_{n,j}(\theta'_n) + \rho \tilde{D}_n \lambda + \varphi_j^*(\tilde{\xi}_{n,j}), \tag{H.91}$$

and

$$\tilde{\mathfrak{w}}_j(\lambda) \equiv \tilde{\mathbb{Z}}_j + \rho D_j \lambda + \pi^*_{1,j}, \tag{H.92}$$

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

$$\begin{aligned} |P_n^*(V_n^I(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset)| &= |\tilde{\mathbf{P}}(\tilde{V}_n^I(\theta'_n, c) \neq \emptyset) - \tilde{\mathbf{P}}(\tilde{\mathfrak{W}}(c) \neq \emptyset)| \\ &\leq |\tilde{\mathbf{P}}(\tilde{V}_n^I(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) + \tilde{\mathbf{P}}(\tilde{V}_n^I(\theta'_n, c) \neq \emptyset \cap \tilde{\mathfrak{W}}(c) = \emptyset)|, \end{aligned} \tag{H.93}$$

where the inequality is due to (H.33). First, we bound the first term on the right hand side of (H.93). Note that

$$\begin{aligned} &\tilde{\mathbf{P}}(\tilde{V}_n^I(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) \\ &\leq \tilde{\mathbf{P}}(\tilde{V}_n^{I,+\delta}(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) + \tilde{\mathbf{P}}(\tilde{V}_n^{I,+\delta}(\theta'_n, c) \neq \emptyset \cap \tilde{V}_n^I(\theta'_n, c) = \emptyset), \end{aligned} \tag{H.94}$$

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

$$\tilde{V}_n^{I,+\delta} \equiv \Big\{\lambda \in B^d_{n,\rho} : p'\lambda = 0 \cap \tilde{v}^I_{n,j,\theta'_n}(\lambda) \leq c+\delta, j \in \mathcal{J}^*\Big\}. \tag{H.95}$$

Let

$$A_n \equiv \Big\{\tilde{\omega} \in \tilde{\Omega} : \sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} |\tilde{v}^I_{n,j,\theta'_n}(\lambda) - \tilde{\mathfrak{w}}_j(\lambda)| \geq \delta\Big\}. \tag{H.96}$$

Let

$$E \equiv \{\{x_i\}_{i=1}^\infty : \|\hat{D}_n(\theta'_n) - D\| < \eta, \max_{j \in \mathcal{J}^*} |\varphi^*_j(\hat{\xi}_{n,j}(\theta'_n)) - \pi^*_{1,j}| < \eta\}. \tag{H.97}$$

Note that, $P_n(E) \geq 1-\eta$ for all $n$ sufficiently large by Assumption E.4 and Lemma H.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 $j \in \mathcal{J}^*$,

$$|\tilde{v}^I_{n,j,\theta'_n}(\lambda) - \tilde{\mathfrak{w}}_j(\lambda)| \leq |\tilde{\mathbb{G}}^b_{n,j}(\theta'_n) - \tilde{\mathbb{Z}}_j| + \rho\|\tilde{D}_{j,n} - D_j\|\|\lambda\| + |\varphi^*_j(\tilde{\xi}_{n,j}) - \pi^*_{1,j}| \leq (2+\rho)\eta, \tag{H.98}$$

uniformly in $\lambda \in B^d$, where we used $\tilde{\mathbb{G}}^b_n \to \tilde{\mathbb{Z}}, \tilde{\mathbf{P}} - a.s.$ Since $\eta$ 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 $\tilde{\mathfrak{W}}(c) \subseteq \tilde{V}_n^{I,+\delta}(\theta'_n, c)$, and hence $A_n^c$ is a subset of

$$L_n \equiv \Big\{\tilde{\omega} \in \tilde{\Omega} : \tilde{\mathfrak{W}}(c) \subseteq \tilde{V}_n^{I,+\delta}(\theta'_n, c)\Big\}. \tag{H.99}$$

Using this,

$$\tilde{\mathbf{P}}(\tilde{V}_n^{I,+\delta}(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) \leq \tilde{\mathbf{P}}(\tilde{\mathfrak{W}}(c) \not\subseteq \tilde{V}_n^{I,+\delta}(\theta'_n, c)) = \tilde{\mathbf{P}}(L_n^c) \leq \tilde{\mathbf{P}}(A_n) < \eta/2, \tag{H.100}$$

for all $n$ sufficiently large. Also, by Lemma H.6,

$$\tilde{\mathbf{P}}(\tilde{V}_n^{I,+\delta}(\theta'_n, c) \neq \emptyset \cap \tilde{V}_n^{I}(\theta'_n, c) = \emptyset) < \eta/2, \tag{H.101}$$

for all $n$ sufficiently large.

Combining (H.94), (H.96), (H.100), (H.101), and using $P_n(E) \geq 1-\eta$ for all $n$, we have

$$\int_E \tilde{\mathbf{P}}(\tilde{V}_n^{I}(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) dP_n + \int_{E^c} \tilde{\mathbf{P}}(\tilde{V}_n^{I}(\theta'_n, c) = \emptyset \cap \tilde{\mathfrak{W}}(c) \neq \emptyset) dP_n \leq \eta(1-\eta) + \eta \leq 2\eta. \tag{H.102}$$

The second term of the right hand side of (H.93) 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

$$P_n^*(V_n^I(\theta'_n, c) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset) \to 0. \tag{H.103}$$

Fix $c > 0$, and set

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j = 1, \ldots, J, \\ 1, & j = J+1, \ldots, J+2d, \\ 0, & j = J+2d+1, J+2d+2. \end{cases} \tag{H.104}$$

Mimic the argument following (H.143). Then, this yields

$$|\Pr(\mathfrak{W}(c) \neq \emptyset) - \Pr(\mathfrak{W}(c-\delta) \neq \emptyset)| = \Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}(c-\delta) = \emptyset\}) \leq \eta, \tag{H.105}$$

$$|\Pr(\mathfrak{W}(c+\delta) \neq \emptyset) - \Pr(\mathfrak{W}(c) \neq \emptyset)| = \Pr(\{\mathfrak{W}(c+\delta) \neq \emptyset\} \cap \{\mathfrak{W}(c) = \emptyset\}) \leq \eta, \tag{H.106}$$

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 $e_j = c - \mathbb{Z}_j$, $j = 1, \ldots, J$. 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

$$K = \begin{bmatrix} D^{[1:J^*,:]} \\ I_d \\ -I_d \\ p' \\ -p' \end{bmatrix}, \; g = \begin{bmatrix} e^{[1:J^*]} \\ \rho \cdot \mathbf{1}_d \\ \rho \cdot \mathbf{1}_d \\ 0 \\ 0 \end{bmatrix}, \; \tau = \begin{bmatrix} \mathbf{1}_{J^*} \\ \mathbf{0}_d \\ \mathbf{0}_d \\ 0 \\ 0 \end{bmatrix}. \tag{H.107}$$

By Farkas' lemma (Rockafellar, 1970, Theorem 22.1) and arguing as in (H.148),

$$\Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{\mathfrak{W}(c-\delta) = \emptyset\}) = \Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\}), \tag{H.108}$$

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 $\{\nu^t \in \mathcal{M}, t = 1, \ldots, T\}$, for which one may write

$$\mathcal{M} = \{\mu : \mu = b \sum_{t=1}^{T} a_t \nu^t, b > 0, a_t \geq 0, \sum_{t=1}^{T} a_t = 1\}. \tag{H.109}$$

This implies

$$\mu' g \geq 0, \; \forall \mu \in \mathcal{M} \;\Leftrightarrow\; \nu^{t\prime} g \geq 0, \; \forall t \in \{1, \ldots, T\} \tag{H.110}$$

$$\mu'(g - \delta\tau) < 0, \; \exists \mu \in \mathcal{M} \;\Leftrightarrow\; \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \; \exists t \in \{1, \ldots, T\}. \tag{H.111}$$

Hence,

$$\begin{aligned} &\Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\}) \\ &\qquad = \Pr\left(0 \leq \nu^{s\prime} g, \; 0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau, \; \forall s, \exists t\right) \end{aligned} \tag{H.112}$$

Note that by (H.107), for each $s \in \{1, \ldots, T\}$,

$$\nu^{s\prime} g = \nu^{s,[1:J^*]\prime}(c 1_{\mathcal{J}^*} - \mathbb{Z}_{\mathcal{J}^*}) + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,[j]}, \tag{H.113}$$

$$\nu^{s\prime} \tau = \sum_{j=1}^{J^*} \nu^{s,[j]}. \tag{H.114}$$

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

$$h_s^U \equiv c \sum_{j=1}^{J^*} \nu^{s,[j]} + \rho \sum_{j=J^*+1}^{J^*+2d} \nu^{s,[j]} \tag{H.115}$$

$$h_s^L \equiv (c-\delta) \sum_{j=1}^{J^*} \nu^{s,[j]}, \tag{H.116}$$

where $0 \le h_s^L < h_s^U$ for all $s \in \{1, \dots, T\}$ 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 (H.112) as

$$\begin{aligned} &\Pr\left(0 \le \nu^{s\prime} g,\ 0 \le \nu^{t\prime} g < \delta \nu^{t\prime}\tau,\ \forall s, \exists t\right) \\ &\qquad = \Pr\left(\nu^{s,[1:J^*]\prime}\mathbb{Z}_{\mathcal{J}^*} \le h_s^U, h_t^L < \nu^{t,[1:J^*]\prime}\mathbb{Z}_{\mathcal{J}^*} \le h_t^U\ \forall s, \exists t\right) > 0, \end{aligned} \tag{H.117}$$

where the last inequality follows because $\mathbb{Z}_{\mathcal{J}^*}$'s correlation matrix $\Omega$ has an eigenvalue bounded away from 0 by Assumption E.3. By (H.108), (H.112), and (H.117), $c \mapsto \Pr(\mathfrak{W}(c) \neq \emptyset)$ 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 $c_n^I(\theta_n') \xrightarrow{P_n} c_{\pi^*}$.

(iii) Begin with observing that one can equivalently express $\hat{c}_n$ (originally defined in (2.13)) as $\hat{c}_n(\theta) = \inf\{c \in \mathbb{R}_+ : P_n^*(V_n^b(\theta, c) \neq \emptyset) \ge 1 - \alpha\}$.

Suppose first that Assumption E.3-1 holds. In this case, there are no paired inequalities, and $V_n^I$ differs from $V_n^b$ only in terms of the function $\varphi_j^*$ in (H.85) used in place of the GMS function $\varphi_j$. In particular, $\varphi_j^*(\xi) \le \varphi_j(\xi)$ for any $j$ and $\xi$, and therefore $\hat{c}_n(\theta_n) \ge c_n^I(\theta_n)$ by construction.

Next, suppose Assumption E.3-2 holds and $V_n^I(\theta_n', c)$ is defined with hard threshold GMS, i.e. with GMS function $\varphi^1$ in AS. The only case that might create concern is one in which

$$\pi_{1,j} \in [-1, 0) \text{ and } \pi_{1,j+R_1} = 0. \tag{H.118}$$

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

$$\pi_{1,j}^* = -\infty, \text{ and } \pi_{1,j+R_1}^* = 0, \tag{H.119}$$

$$\varphi_j(\hat{\xi}_{n,j}(\theta_n')) = 0, \text{ and } \varphi_{j+R_1}(\hat{\xi}_{n,j+R_1}(\theta_n')) = 0, \tag{H.120}$$

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

$$\mathbb{G}_{n,j+R_1}^b(\theta_n') + \rho \hat{D}_{n,j+R_1}(\theta_n')\lambda \le c, \tag{H.121}$$

is replaced with inequality

$$-\mathbb{G}_{n,j}^b(\theta_n') - \rho \hat{D}_{n,j}(\theta_n')\lambda \le c, \tag{H.122}$$

as explained in Section E.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 (H.83) but replacing $j + R_1$-th component $\mathbb{G}_{n,j+R_1}^b(\theta_n) + \hat{D}_{n,j+R_1}(\theta_n)\lambda +$

$\varphi^*_{j+R_1}(\hat{\xi}_{n,j+R_1}(\theta_n))$ 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 (H.86) but replacing $v_n^I$ with $v_n^{IP}$. Define $c_n^{IP}(\theta_n) \equiv \inf\{c \in \mathbb{R}_+ : P^*(V_n^{IP}(\theta_n, c)) \geq 1-\alpha\}$. By construction, $\hat{c}_n(\theta'_n) \geq c_n^{IP}(\theta'_n)$ for any $\theta'_n \in (\theta_n + \rho/\sqrt{n}B^d) \cap \Theta$. Therefore, it suffices to show that $c_n^{IP}(\theta'_n) - c_n^I(\theta'_n) \stackrel{P_n}{\to} 0$. For this, note that Lemma H.9-(3) establishes

$$\sup_{\lambda \in B^d_{n,\rho}} \|\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda + \mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\| = o_{P^*}(1), \tag{H.123}$$

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

$$c_n^{IP}(\theta'_n) \stackrel{P_n}{\to} c_{\pi^*}. \tag{H.124}$$

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

If the set $V_n^I(\theta'_n, c)$ is defined with a GMS function satisfying Assumption E.2 and continuous in its argument, we can mimic the above argument using the replacements in (H.12)-(H.13) with $\hat{\mu}_{n,j+R_1}$ as defined in (H.14) and $\hat{\mu}_{n,j}(\theta'_n)$ as in (H.15). Then when both $\pi_j \in (-\infty, 0]$ and $\pi_{j+R_1} \in (-\infty, 0]$ we have:

$$\begin{aligned}\Delta(\mu,\hat{\mu}) \equiv \Big\|&\hat{\mu}_{n,j}(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\} - \hat{\mu}_{n,j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\}\\ &-\mu_j(\theta'_n)\{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\} + \mu_{j+R_1}(\theta'_n)\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\}\Big\|\\ &= o_{\mathcal{P}}(1),\end{aligned} \tag{H.125}$$

where $\mu_j, \mu_{j+R_1}$ are defined in equations (H.10)-(H.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

$$\begin{aligned}\Delta(\mu,\hat{\mu}) \leq &\left|\hat{\mu}_{n,j+R_1}(\theta'_n) - \mu_{j+R_1}(\theta'_n)\right|\\ &\times \left\|\{\mathbb{G}^b_{n,j+R_1}(\theta'_n) + \rho\hat{D}_{n,j+R_1}(\theta'_n)\lambda\} + \{\mathbb{G}^b_{n,j}(\theta'_n) + \rho\hat{D}_{n,j}(\theta'_n)\lambda\}\right\|.\end{aligned} \tag{H.126}$$

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 H.9-(3)-(4), and by Assumption E.4-(i). The rest of the argument follows similarly as for the case of hard-threshold GMS. *Q.E.D.*

LEMMA H.4 *Let Assumptions E.1, E.2, E.4, and E.5 hold. Let $(P_n, \theta_n)$ be the sequence satisfying (H.1)-(H.3), let $\mathcal{J}^*$ be defined as in (H.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 $n \geq \max\{N', N''\}$,*

$$\mathbf{P}\left(\sup_{\lambda \in B^d}\left|\max_{j=1,\dots,J}(u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j=1,\dots,J}(\mathfrak{w}^*_j(\lambda) - c_{\pi^*})\right| \geq \varepsilon\right) < \eta, \tag{H.127}$$

$$\tilde{\mathbf{P}}\left(\sup_{\lambda \in B^d}\left|\max_{j=1,\dots,J}\tilde{\mathfrak{w}}_j(\lambda) - \max_{j=1,\dots,J}\tilde{v}^I_{n,j,\theta'_n(\lambda)}\right| \geq \varepsilon\right) < \eta,\ w.p.1, \tag{H.128}$$

*where the functions $u^*_n, \mathfrak{w}^*, \tilde{v}_n, \tilde{\mathfrak{w}}$ are defined in equations (H.24),(H.25), (H.91), and (H.92).*

PROOF: We first establish (H.127). By definition, $\pi^*_{1,j} = -\infty$ for all $j \notin \mathcal{J}^*$ and therefore

$$\mathbf{P}\Big( \sup_{\lambda \in B^d} | \max_{j=1,\dots,J} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j=1,\dots,J} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \geq \varepsilon \Big) \tag{H.129}$$

$$= \mathbf{P}\Big( \sup_{\lambda \in B^d} | \max_{j \in \mathcal{J}^*} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j \in \mathcal{J}^*} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \geq \varepsilon \Big). \tag{H.130}$$

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

$$\begin{aligned} |r_{n,j,\theta_n}(\lambda)| &\leq |\mathbb{G}^*_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j| + \rho \|D_{P_n,j}(\bar\theta_n) - D_j\| \|\lambda\| + |\mathbb{G}^*_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) \\ &\qquad + \rho D_{P_n,j}(\bar\theta_n)\lambda| \; \eta^*_{n,j} + |c^*_n - c_{\pi^*}| \\ &= |\mathbb{G}^*_{n,j}(\theta_n + \tfrac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j| + o(1) + \{O_{\mathcal{P}}(1) + O(1)\})\eta^*_{n,j} + o_{\mathcal{P}}(1) \end{aligned} \tag{H.131}$$

$$= o_{\mathcal{P}}(1) \tag{H.132}$$

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 E.4-(ii), and $\bar\theta_n$ being a mean value between $\theta_n$ 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 E.4-(i)), and $c^*_n \stackrel{a.s.}{\to} c_{\pi^*}$. The last equality follows from $\mathbb{G}^*_{n,j}(\theta_n + \frac{\lambda\rho}{\sqrt{n}}) - \mathbb{Z}^*_j \stackrel{a.s.}{\to} 0$ and $\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)| = o_{\mathcal{P}}(1)$ by Lemma H.10.

We note that when paired inequalities are merged, 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 $\mu_j$ were defined in (G.11)-(G.12) and (H.10)-(H.11) respectively.

By (H.132) and the fact that $j \in \mathcal{J}^*$, we have

$$\sup_{\lambda \in B^d} | \max_{j \in \mathcal{J}^*} (u^*_{n,j,\theta_n}(\lambda) - c^*_n) - \max_{j \in \mathcal{J}^*} (\mathfrak{w}^*_j(\lambda) - c_{\pi^*})| \leq \sup_{\lambda \in B^d} \max_{j \in \mathcal{J}^*} |r_{n,j,\theta_n}(\lambda)| = o_{\mathcal{P}}(1). \tag{H.133}$$

The conclusion of the lemma then follows from (H.130) and (H.133).

The result in (H.128) follows from similar arguments. *Q.E.D.*

LEMMA H.5 *Let Assumptions E.1, E.2, E.4, and E.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 $j = 1, \dots, J$, let $\chi_j(\{Q_n, \vartheta_n\})$ be a function of the sequence $\{Q_n, \vartheta_n\}$ defined as*

$$\chi_j(\{Q_n, \vartheta_n\}) \equiv \begin{cases} 0, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\vartheta_n) = 0, \\ -\infty, & \text{if } \lim_{n\to\infty} \kappa_n^{-1}\sqrt{n}\gamma_{1,Q_n,j}(\vartheta_n) < 0. \end{cases} \tag{H.134}$$

*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

$$\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P),\theta'\in\theta+\rho/\sqrt{n}B^d}\left|\frac{\sqrt{n}E_P(m_j(X,\theta))}{\kappa_n\sigma_{P,j}(\theta)}-\frac{\sqrt{n}E_P(m_j(X,\theta'))}{\kappa_n\sigma_{P,j}(\theta')}\right| \le \sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P),\theta'\in\theta+\rho/\sqrt{n}B^d}\frac{\sqrt{n}\|D_{P,j}(\tilde{\theta})\|\|\theta'-\theta\|}{\kappa_n}=o(1), \tag{H.135}$$

where $\tilde{\theta}$ represents a mean value that lies componentwise between $\theta$ and $\theta'$ and where we used the fact that $D_{P,j}(\theta)$ is Lipschitz continuous and $\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P)}\|D_{P,j}(\theta)\|\le\bar{M}$. Result (ii) then follows immediately from (H.134). For (iii), note that

$$\begin{aligned}&\sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}\left|\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n')}{\hat{\sigma}_{n,j}(\theta_n')}-\kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\right|\\&\le\sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}\left|\kappa_n^{-1}\frac{\sqrt{n}(\bar{m}_{n,j}(\theta_n')-E_{P_n}[m_j(X_i,\theta_n')])}{\sigma_{n,j}(\theta_n')}(1+\eta_{n,j}(\theta_n'))\right.\\&\quad\left.+\kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\right|\\&\le\sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}|\kappa_n^{-1}\mathbb{G}_n(\theta_n')(1+\eta_{n,j}(\theta_n'))|+\left|\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\kappa_n\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\right|=o_{\mathcal{P}}(1),\end{aligned} \tag{H.136}$$

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 (2015b), 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 H.10-(i). *Q.E.D.*

LEMMA H.6 *Let Assumptions E.1, E.2, E.3, E.4, and E.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*

$$\sup_{c\ge0}\Pr(\{\mathfrak{W}(c)\neq\emptyset\}\cap\{\mathfrak{W}^{-\delta}(c)=\emptyset\})<\eta. \tag{H.137}$$

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

$$\sup_{c\ge0}P_n^*(\{V_n^I(\theta_n',c)\neq\emptyset\}\cap\{V_n^{I,-\delta}(\theta_n',c)=\emptyset\})<\eta,\ \forall n\ge N. \tag{H.138}$$

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

$$\mathfrak{W}^{-\delta}(c)\equiv\left\{\lambda\in\mathfrak{B}_\rho^d:p'\lambda=0\cap\mathfrak{w}_j(\lambda)\le c-\delta,\ \forall j=1,\dots,J\right\}, \tag{H.139}$$

*and*

$$V_n^{I,-\delta}(\theta_n',c)\equiv\left\{\lambda\in B_{n,\rho}^d:p'\lambda=0\cap v_{n,j,\theta_n'}^I(\lambda)\le c-\delta,\ \forall j=1,\dots,J\right\}. \tag{H.140}$$

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

$$\sup_{c\ge\underline{c}}\Pr(\{\mathfrak{W}(c)\neq\emptyset\}\cap\{\mathfrak{W}^{-\delta}(c)=\emptyset\})<\eta. \tag{H.141}$$

*with* $\mathfrak{W}^{-\delta}(c)$ *defined in* (H.139). *Moreover, for any* $\eta > 0$, *there exist* $\delta > 0$ *and* $N \in \mathbb{N}$ *such that*

$$\sup_{c \geq \underline{c}} P_n^*(\{V_n^I(\theta_n', c) \neq \emptyset\} \cap \{V_n^{I,-\delta}(\theta_n', c) = \emptyset\}) < \eta, \ \forall n \geq N, \tag{H.142}$$

*with* $V_n^{-\delta}(\theta_n', c)$ *defined in* (H.140).

PROOF: We first show (H.137). If $\mathcal{J}^* = \emptyset$, with $\mathcal{J}^*$ as defined in (H.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 (G.21) (both with $c$ and with $c - \delta$). Hence, one could argue for sets $\mathfrak{W}(c), \mathfrak{W}^{-\delta}(c)$ defined as in equations (H.16) and (H.17) but with $j \in \mathcal{J}^*$. To keep the notation simple, below we argue as if all $j = 1, \dots, J$ belong to $\mathcal{J}^*$. Let $c \geq 0$ be given. Let $g$ be a $J + 2d + 2$ vector with entries

$$g_j = \begin{cases} c - \mathbb{Z}_j, & j = 1, \dots, J, \\ 1, & j = J + 1, \dots, J + 2d, \\ 0, & j = J + 2d + 1, J + 2d + 2, \end{cases} \tag{H.143}$$

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

$$\tau_j = \begin{cases} 1, & j = 1, \dots, J_1, \\ 0, & j = J_1 + 1, \dots, J + 2d + 2. \end{cases} \tag{H.144}$$

Then we can express the sets of interest as

$$\mathfrak{W}(c) = \{\lambda : K\lambda \leq g\}, \tag{H.145}$$

$$\mathfrak{W}^{-\delta}(c) = \{\lambda : K\lambda \leq g - \delta\tau\}. \tag{H.146}$$

By Farkas' Lemma, e.g. Rockafellar (1970, Theorem 22.1), a solution to the system of linear inequalities in (H.145) 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 (H.146) 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

$$\mathcal{M} \equiv \{\mu \in \mathbb{R}_+^{J+2d+2} : \mu' K = 0\}. \tag{H.147}$$

Then, one may write

$$\begin{aligned} &\Pr(\{\mathfrak{W}(c) \neq \emptyset\} \cap \{W^{-\delta}(\theta_n', c) = \emptyset\}) \\ &= \Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu'(g - \delta\tau) < 0, \exists \mu \in \mathcal{M}\}) \\ &= \Pr(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu' g < \delta\mu'\tau, \exists \mu \in \mathcal{M}\}). \end{aligned} \tag{H.148}$$

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 $\{\nu^t \in \mathcal{M}, t = 1, \dots, T\}$, with $T < \infty$ a constant that depends only on $J$ and $d$, such that any $\mu \in \mathcal{M}$ can be represented as

$$\mu = b \sum_{t=1}^{T} a_t \nu^t, \tag{H.149}$$

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 $\nu^t$, we have

$$\sum_{t=1}^{T} a_t \nu^{t\prime} g < \delta \sum_{t=1}^{T} a_t \nu^{t\prime} \tau. \tag{H.150}$$

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,

$$\begin{aligned} &\Pr\left(\{\mu' g \geq 0, \forall \mu \in \mathcal{M}\} \cap \{\mu' g < \delta \mu' \tau, \exists \mu \in \mathcal{M}\}\right) \\ &\qquad \leq \Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau, \exists t \in \{1, \ldots, T\}\right) \leq \sum_{t=1}^{T} \Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right). \end{aligned} \tag{H.151}$$

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

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

where the last equality follows by (H.144). Therefore $\Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime} \tau\right) = 0$.
**Case 2.** Consider now any $t = 1, \ldots, T$ such that $\nu^t$ 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:

$$g = \begin{cases} c - \mathbb{Z}_j & j = J_1 + 1, \ldots, J_1 + J_2, \\ c + \mathbb{Z}_{j-J_2} & j = J_1 + J_2 + 1, \ldots, J. \end{cases} \tag{H.152}$$

For each $\nu^t$, (H.152) implies

$$\sum_{j=J_1+1}^{J_1+2J_2} \nu_j^t g_j = c \sum_{j=J_1+1}^{J_1+2J_2} \nu_j^t + \sum_{j=J_1+1}^{J_1+J_2} (\nu_j^t - \nu_{j+J_2}^t) \mathbb{Z}_j. \tag{H.153}$$

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

$$\tilde{\nu}_j^t \equiv \begin{cases} \nu_j^t & j = 1, \ldots, J_1 \\ \nu_j^t - \nu_{j+J_2}^t & j = J_1 + 1, \ldots, J_1 + J_2. \end{cases} \tag{H.154}$$

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

$$\nu^{t\prime} g = \sum_{j=1}^{J_1+J_2} \tilde{\nu}_j^t \mathbb{Z}_j + c \sum_{j=1}^{J} \nu_j^t + \sum_{j=J+1}^{J+2d} \nu_j^t. \tag{H.155}$$

**Case 2-a.** Suppose $\tilde{\nu}^t \neq 0$. Then, by (H.155), $\frac{\nu^{t\prime} g}{\nu^{t\prime}\tau}$ is a normal random variable with variance $(\tilde{\nu}^{t\prime}\tau)^{-2}\tilde{\nu}^{\prime t}\Omega\tilde{\nu}^t$. By Assumption E.3, there exists a constant $\omega > 0$ such that the smallest eigenvalue of $\Omega$ is bounded from below by $\omega$ for all $\theta_n'$. Hence, letting $\|\cdot\|_p$ denote the $p$-norm in $\mathbb{R}^{J+2d+2}$, we have

$$\frac{\tilde{\nu}^{\prime t}\Omega\tilde{\nu}^t}{(\tilde{\nu}^{t\prime}\tau)^2} \geq \frac{\omega\|\tilde{\nu}^t\|_2^2}{(J+2d+2)^2\|\tilde{\nu}^t\|_2^2} \geq \frac{\omega}{(J+2d+2)^2}. \tag{H.156}$$

Therefore, the variance of the normal random variable in (H.151) 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 $\nu^t$ assigns positive weight also to constraints in $\{1,\ldots,J\}$, this must be because $\nu_j^t = 0$ for all $j = 1,\ldots,J_1$ and $\nu_j^t = \nu_{j+J_2}^t$ for all $j = J_1+1,\ldots,J_1+J_2$, while $\nu_j^t \neq 0$ for some $j = J_1+1,\ldots,J_1+J_2$. 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 $\nu^t$ assigns positive weight only to constraints in $\{J+1,\ldots,J+2d+2\}$ applies and again $\Pr\left(0 \leq \nu^{t\prime} g < \delta \nu^{t\prime}\tau\right) = 0$. This establishes equation (H.137).

To see why equation (H.138) 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 (H.58) 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 (H.147) 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 (2015b) together with Lemma H.17 (through an argument similar to that following equation (H.90)), $\mathbb{G}_n^b \xrightarrow{d} \mathbb{G}_P$ in $l^\infty(\Theta)$ uniformly in $\mathcal{P}$ conditional on $\{X_1,\ldots,X_n\}$, and by Assumption E.4 $\hat{D}_n(\theta_n') \xrightarrow{P_n} D$, for almost all sample paths. Set

$$g_{P_n,j}(\theta_n') = \begin{cases} c - \varphi_j^*(\xi_{n,j}(\theta_n')) - \mathbb{G}_{n,j}^b(\theta_n'), & j = 1,\ldots,J, \\ 1, & j = J+1,\ldots,J+2d, \\ 0, & j = J+2d+1, J+2d+2, \end{cases} \tag{H.157}$$

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 (H.143) to conclude (H.138).

The results in (H.141)-(H.142) follow by similar arguments, with proper redefinition of $\tau$ in equation (H.144). *Q.E.D.*

LEMMA H.7 *Let Assumptions E.3 and E.5 hold. Let $(P_n,\theta_n)$ have the almost sure representations given in Lemma H.1, let $\mathcal{J}^*$ be defined as in (H.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 H.2. Then, for any $\eta > 0$, there exist $0 < \varepsilon_\eta < \infty$ and $N \in \mathbb{N}$ s.t. $n \geq N$ implies*

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \eta. \tag{H.158}$$

PROOF: We bound the probability in (H.158) as follows:

$$\mathbf{P}\left(\mathfrak{W}^{*,-\delta}(0) \neq \emptyset, \left|\det K^{\mathcal{C}}\right| \leq \varepsilon_\eta\right) \leq \mathbf{P}\left(\exists C \in \widetilde{\mathcal{C}} : \lambda^C \in B^d, \left|\det K^C\right| \leq \varepsilon_\eta\right) \tag{H.159}$$

$$\leq \sum_{C \in \widetilde{\mathcal{C}}: \left|\det K^C\right| \leq \varepsilon_\eta} \mathbf{P}\left(\lambda^C \in B^d\right) \tag{H.160}$$

$$\leq \sum_{C \in \widetilde{\mathcal{C}}: |\alpha^C| \leq \varepsilon_\eta^{2/d}} \mathbf{P}\left(\lambda^C \in B^d\right), \tag{H.161}$$

where $\alpha^C$ 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 $\alpha^C$ and recall that because $K^C K^{C\prime}$ is symmetric, $\|q^C\| = 1$. Thus the claim is equivalent to:

$$|q^{C\prime} K^C K^{C\prime} q^C| \leq \varepsilon_\eta^{2/d} \implies \mathbf{P}((K^C)^{-1} g^C \in \mathfrak{B}_\rho^d) \leq \overline{\eta}. \tag{H.162}$$

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

$$\left|q^{C\prime} g_{P_n}^C\right| = \left|q^{C\prime} K^C \left(K^C\right)^{-1} g^C\right| < \sqrt{d}\varepsilon_\eta^{1/d}, \tag{H.163}$$

hence

$$\mathbf{P}((K^C)^{-1} g^C \in \mathfrak{B}_\rho^d) \leq \mathbf{P}\Big(|q^{C\prime} g^C| < \sqrt{d}\varepsilon_\eta^{1/d}\Big). \tag{H.164}$$

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 E.3 and E.5 (iii) yield

$$\begin{aligned} & \mathrm{eig}(\tilde{\Omega}) \geq \omega \\ \Longrightarrow\ & Var_{\mathbf{P}}(q^{C\prime} g^C) \geq \omega \\ \Longrightarrow\ & \mathbf{P}\left(|q^{C\prime} g^C| < \sqrt{d}\varepsilon_\eta^{1/d}\right) = \mathbf{P}\left(-\sqrt{d}\varepsilon_\eta^{1/d} < q^{C\prime} g^C < \sqrt{d}\varepsilon_\eta^{1/d}\right) \\ & < \frac{2\sqrt{d}\varepsilon_\eta^{1/d}}{\sqrt{2\omega\pi}}, \end{aligned} \tag{H.165}$$

where the result in (H.165) uses that the density of a normal r.v. is maximized at the expected value. The result follows by choosing

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

*Q.E.D.*

LEMMA H.8 *Let Assumptions E.1, E.2, E.3, E.4, and E.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 $c \geq 0$ 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 $\theta$. Also, if $C$ is an index set picking certain constraints, then $\hat{D}_n^C$ is the matrix collecting the corresponding estimated gradients, and similarly for $\mathbb{G}_n^{b,C}$.

Suppose now case (i), then there exists an index set $\bar{C} \subset \{J_1+1, ..., J_1+d, J_1+J_2+1, \ldots, J_1+J_2+d\}$ picking one direction of each constraint s.t. $p$ is a positive linear combination of the rows of $\hat{D}_P^{\bar{C}}$. (This choice ensures that a Karush-Kuhn-Tucker condition holds, justifying the step from (H.167) to (H.168) below.) Then the coverage probability $P^*(V_n^I(\theta, c) \neq \emptyset)$ is asymptotically bounded above by

$$P^*\Big(\sup_{\lambda\in\rho B^d_{n,\rho}} \Big\{p'\lambda : \hat{D}_{n,j}\lambda \leq c - \mathbb{G}^b_{n,j}, j \in \mathcal{J}^*\Big\} \geq 0\Big) \tag{H.166}$$

$$\leq P^*\Big(\sup_{\lambda\in\mathbb{R}^d} \Big\{p'\lambda : \hat{D}_{n,j}\lambda \leq c - \mathbb{G}^b_{n,j}, j \in \bar{C}\Big\} \geq 0\Big) \tag{H.167}$$

$$= P^*\Big(p'(\hat{D}_n^{\bar{C}})^{-1}(c\mathbf{1}_d - \mathbb{G}_n^{b,\bar{C}}) \geq 0\Big) \tag{H.168}$$

$$= P^*\left(\frac{p'(\hat{D}_n^{\bar{C}})^{-1}(c\mathbf{1}_d - \mathbb{G}_n^{b,\bar{C}})}{\sqrt{p'(\hat{D}_n^{\bar{C}})^{-1}\Omega_P^C(\hat{D}_n^{\bar{C}})^{-1}p}} \geq 0\right) \tag{H.169}$$

$$= P^*\left(\frac{p'adj(\hat{D}_n^{\bar{C}})(c\mathbf{1}_d - \mathbb{G}_n^{b,\bar{C}})}{\sqrt{p'(adj(\hat{D}_n^{\bar{C}})\Omega_P^C adj(\hat{D}_n^{\bar{C}})p}} \geq 0\right) \tag{H.170}$$

$$= \Phi\left(\frac{p'adj(\hat{D}_n^{\bar{C}})c\mathbf{1}_d}{\sqrt{p'(adj(\hat{D}_n^{\bar{C}})\Omega_P^C adj(\hat{D}_n^{\bar{C}})p}}\right) + o_{\mathcal{P}}(1) \tag{H.171}$$

$$\leq \Phi(d\omega^{-1/2}c) + o_{\mathcal{P}}(1). \tag{H.172}$$

Here, (H.167) removes constraints and hence enlarges the feasible set; (H.168) solves in closed form; (H.169) divides through by a positive scalar; (H.170) eliminates the determinant of $\hat{D}_n^{\bar{C}}$, using that rows of $\hat{D}_n^{\bar{C}}$ can always be rearranged so that the determinant is positive; (H.171) follows by Assumption E.5, using that the term multiplying $\mathbb{G}_n^{b,\bar{C}}$ is $O_{\mathcal{P}}(1)$; and (H.172) uses that by Assumption E.3, there exists a constant $\omega > 0$ that does not depend on $\theta$ such that the smallest eigenvalue of $\Omega_P$ is bounded from below by $\omega$. 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 $\hat{D}_P^{\bar{C}}$. (Note that $\bar{C}$ cannot contain $J_1+1$ or $J_1+J_2+1$.) One can then write

$$P^*\Big(\sup_{\lambda\in\rho B^d_{n,\rho}}\Big\{p'\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\cup\{J_1+J_2+1\}\Big\}\ge 0\Big) \tag{H.173}$$

$$\le P^*\left(\exists\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\cup\{J_1+J_2+1\}\right) \tag{H.174}$$

$$\le P^*\Bigg(\sup_{\lambda\in\rho B^d_{n,\rho}}\left\{\hat{D}_{n,J_1+1}\lambda : \hat{D}_{n,j}\lambda \le c - \mathbb{G}^b_{n,j}, j\in\bar{C}\right\} \tag{H.175}$$

$$\ge \inf_{\lambda\in\rho B^d_{n,\rho}}\left\{\hat{D}_{n,J_1+1}\lambda : \hat{D}_{n,J_1+J_2+1}\lambda \le c - \mathbb{G}^b_{n,J_1+J_2+1}\right\}\Bigg) \tag{H.176}$$

$$= P^*\left(\hat{D}_{n,J_1+1}\hat{D}_n^{\bar{C}\prime}(\hat{D}_n^{\bar{C}}\hat{D}_n^{\bar{C}\prime})^{-1}(c\mathbf{1}_{\bar{d}} - \mathbb{G}_n^{b,\bar{C}}) \ge -c + \mathbb{G}^b_{n,J_1+J_2+1}\right). \tag{H.177}$$

Here, the reasoning from (H.173) to (H.176) holds because we evaluate the probability of increasingly larger events; in particular, if the event in (H.176) 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 (H.177) resembles (H.169), and the argument can be concluded in analogy to (H.170)-(H.172). *Q.E.D.*

LEMMA H.9 *Let Assumptions E.1, E.2, E.3-2, E.4, and E.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 (G.4). Let $(P_n,\theta_n)$ be the sequence satisfying the conditions of Lemma H.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 $j=1,\dots,R_1$.*
*(2) $Corr_{P_n}(m_j(X_i,\theta'_n), m_{j+R_1}(X_i,\theta'_n))\to -1$ for $j=1,\dots,R_1$.*
*(3) $|\mathbb{G}_{n,j}(\theta'_n)+\mathbb{G}_{n,j+R_1}(\theta'_n)|\stackrel{P_n}{\to}0$, and $|\mathbb{G}^b_{n,j}(\theta'_n)+\mathbb{G}^b_{n,j+R_1}(\theta'_n)|\stackrel{P^*_n}{\to}0$ 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 H.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.**

$$\lim_{n\to\infty}\frac{\kappa_n}{\sqrt{n}}\sigma_{P_n,j}(\theta'_n)>0, \tag{H.178}$$

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

$$\frac{\sigma^2_{P_n,j+R_1}(\theta'_n)}{\sigma^2_{P_n,j}(\theta'_n)} = \frac{\sigma^2_{P_n,j}(\theta'_n)+Var_{P_n}(t_j(X_i,\theta'_n))+2Cov_{P_n}(m_j(X_i,\theta'_n),t_j(X_i,\theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)}\to 1, \tag{H.179}$$

where the convergence follows because $Var_{P_n}(t_j(X_i,\theta'_n))$ is bounded due to Assumption E.3-2,

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

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

$$\lim_{n\to\infty} \frac{\kappa_n}{\sqrt{n}} \sigma_{P_n,j}(\theta'_n) = 0. \tag{H.180}$$

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 (H.179), it also follows that

$$\lim_{n\to\infty} \frac{\kappa_n}{\sqrt{n}} \sigma_{P_n,j+R_1}(\theta'_n) = 0, \tag{H.181}$$

and hence that $E_{P_n}(t_j(X_i, \theta'_n)) \to 0$. We then have, using Assumption E.3-2 again,

$$\begin{aligned} Var_{P_n}(t_j(X_i, \theta'_n)) &= \int t_j(x, \theta'_n)^2 dP_n(x) - E_{P_n}[t_j(X_i, \theta'_n)]^2 \\ &\leq M \int t_j(x, \theta'_n) dP_n(x) - E_{P_n}[t_j(X_i, \theta'_n)]^2 \to 0. \end{aligned} \tag{H.182}$$

Hence,

$$\begin{aligned} \frac{\sigma^2_{P_n,j+R_1}(\theta'_n)}{\sigma^2_{P_n,j}(\theta'_n)} &= \frac{\sigma^2_{P_n,j}(\theta'_n) + Var_{P_n}(t_j(X_i, \theta'_n)) + 2Cov_{P_n}(m_j(X_i, \theta'_n), t_j(X_i, \theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)} \\ &\leq \frac{\sigma^2_{P_n,j}(\theta'_n) + Var_{P_n}(t_j(X_i, \theta'_n))}{\sigma^2_{P_n,j}(\theta'_n)} + \frac{2(Var_{P_n}(t_j(X_i, \theta'_n)))^{1/2}}{\sigma_{P_n,j}(\theta'_n)} \\ &\to 1, \end{aligned} \tag{H.183}$$

and the first claim follows.
To obtain claim 2, note that

$$\begin{aligned} Corr_{P_n}(m_j(X_i, \theta'_n), m_{j+R_1}(X_i, \theta'_n)) &= \frac{-\sigma^2_{P_n,j}(\theta'_n) - Cov_{P_n}(m_j(X_i, \theta'_n), t_j(X_i, \theta'_n))}{\sigma_{P_n,j}(\theta'_n)\sigma_{P_n,j+R_1}(\theta'_n)} \\ &\to -1, \end{aligned} \tag{H.184}$$

where the result follows from (H.182) and (H.183).
To establish Claim 3, consider $\mathbb{G}_n$ below. Note that, for $j = 1, \dots, R_1$,

$$\begin{bmatrix} \mathbb{G}_{n,j}(\theta'_n) \\ \mathbb{G}_{n,j+R_1}(\theta'_n) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (m_j(X_i,\theta'_n) - E_{P_n}[m_j(X_i,\theta'_n)])}{\sigma_{P_n,j}(\theta'_n)} \\ -\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (m_j(X_i,\theta'_n) - E_{P_n}[m_j(X_i,\theta'_n)]) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (t_j(X_i,\theta'_n) - E_{P_n}[t_j(X_i,\theta'_n)])}{\sigma_{P_n,j+R_1}(\theta'_n)} \end{bmatrix}. \tag{H.185}$$

Under the conditions of Case 1 above, we immediately obtain

$$|\mathbb{G}_{n,j}(\theta'_n) + \mathbb{G}_{n,j+R_1}(\theta'_n)| \xrightarrow{P_n} 0. \tag{H.186}$$

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, (H.186) obtains again. These results imply that $\mathbb{Z}_j + \mathbb{Z}_{j+R_1} = 0, a.s.$ By Lemma H.15, $\{\mathbb{G}^b_n\}$

converges in law to the same limit as $\{\mathbb{G}_n\}$ for almost all sample paths $\{X_i\}_{i=1}^{\infty}$. This and (H.186) 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

$$E_{P_n}\left(\frac{m_j(X,\theta'_n)}{\sigma_{P_n,j}(\theta'_n)}+\frac{m_{j+R_1}(X,\theta'_n)}{\sigma_{P_n,j+R_1}(\theta'_n)}\right)=O_{\mathcal{P}}\left(\frac{\kappa_n}{\sqrt{n}}\right). \tag{H.187}$$

Define the $1\times d$ vector

$$q_n\equiv D_{P_n,j+R_1}(\theta'_n)+D_{P_n,j}(\theta'_n). \tag{H.188}$$

Suppose by contradiction that

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

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

$$\tilde r_n=\frac{q'_n}{\|q_n\|}. \tag{H.189}$$

Let

$$r_n=\tilde r_n\rho\kappa_n^2/\sqrt{n}. \tag{H.190}$$

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 $\theta'_n$ and $\theta'_n+r_n$, we obtain

$$\begin{aligned}
&E_{P_n}\Big(\frac{m_j(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}+\frac{m_{j+R_1}(X,\theta'_n+r_n)}{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}\Big)\\
&\quad=E_{P_n}\left(\frac{m_j(X,\theta'_n)}{\sigma_{P_n,j}(\theta'_n)}+\frac{m_{j+R_1}(X,\theta'_n)}{\sigma_{P_n,j+R_1}(\theta'_n)}\right)+\big(D_{P_n,j}(\bar\theta_n)+D_{P_n,j+R_1}(\tilde\theta_n)\big)r_n\\
&\quad=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+(D_{P_n,j}(\theta'_n)+D_{P_n,j+R_1}(\theta'_n))\,r_n+\big(D_{P_n,j}(\bar\theta_n)-D_{P_n,j}(\theta'_n)\big)\,r_n\\
&\qquad+\big(D_{P_n,j+R_1}(\tilde\theta_n)-D_{P_n,j+R_1}(\theta'_n)\big)r_n\\
&\quad=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+\frac{\rho\kappa_n^2}{\sqrt{n}}+O_{\mathcal{P}}(\frac{\rho^2\kappa_n^4}{n}).
\end{aligned}\tag{H.191}$$

It then follows that there exists $N\in\mathbb{N}$ such that for all $n\geq N$, the right hand side in (H.191) is strictly greater than zero.

Next, observe that

$$\begin{aligned}
&E_{P_n}\Big(\frac{m_j(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}+\frac{m_{j+R_1}(X,\theta'_n+r_n)}{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}\Big)\\
&\quad=E_{P_n}\left(\frac{m_j(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}+\frac{m_{j+R_1}(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}\right)\\
&\qquad-\left(\frac{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}-1\right)\frac{E_{P_n}(m_{j+R_1}(X,\theta'_n+r_n))}{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}\\
&\quad=E_{P_n}\left(\frac{m_j(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}+\frac{m_{j+R_1}(X,\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}\right)-o_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}}).
\end{aligned}\tag{H.192}$$

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 E.3-2, we have

$$\frac{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}{\sigma_{P_n,j}(\theta'_n+r_n)}-1=\frac{\sigma_{P_n,j+R_1}(\theta'_n)}{\sigma_{P_n,j}(\theta'_n)}-1+o_{\mathcal{P}}(1)=o_{\mathcal{P}}(1), \tag{H.193}$$

where we used Claim 1. Second, using Assumption E.4, we have that

$$\begin{aligned}\frac{E_{P_n}(m_{j+R_1}(X,\theta'_n+r_n))}{\sigma_{P_n,j+R_1}(\theta'_n+r_n)}&=\frac{E_{P_n}(m_{j+R_1}(X,\theta'_n))}{\sigma_{P_n,j+R_1}(\theta'_n)}+D_{P_n,j+R_1}(\tilde{\theta}_n)r_n\\&=O_{\mathcal{P}}(\frac{\kappa_n}{\sqrt{n}})+O_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}}).\end{aligned} \tag{H.194}$$

The product of (H.193) and (H.194) is therefore $o_{\mathcal{P}}(\frac{\rho\kappa_n^2}{\sqrt{n}})$ and (H.192) follows.

To conclude the argument, note that for $n$ large enough, $m_{j+R_1}(X,\theta'_n+r_n)\le -m_j(X,\theta'_n+r_n)$ *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 E.3-2 applies. Therefore, there exists $N\in\mathbb{N}$ such that for all $n\ge N$, the left hand side in (H.191) is strictly less than the right hand side, yielding a contradiction. *Q.E.D.*

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

LEMMA H.10 *Suppose Assumptions E.1, E.2, and E.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 $j=1,\dots,J_1+J_2$*

$$\inf_{P\in\mathcal{P}}P\Big(\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)|\to 0\Big)=1. \tag{H.195}$$

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

$$\hat{\sigma}^M_{n,j}(\theta)=\hat{\sigma}^M_{n,j+R_1}(\theta)\equiv\hat{\mu}_{n,j}(\theta)\hat{\sigma}_{n,j}(\theta)+(1-\hat{\mu}_{n,j}(\theta))\hat{\sigma}_{n,j+R_1}(\theta). \tag{H.196}$$

*Let $(P_n,\theta_n)$ 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 (H.29). Then, for any $\eta>0$, there exists $N\in\mathbb{N}$ such that*

$$P_n\Big(\max_{j\in(\mathcal{R}_1\cup\mathcal{R}_2)\cap\mathcal{J}^*}\Big|\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}-1\Big|>\eta\Big)<\eta \tag{H.197}$$

*for all $n\ge N$.*

PROOF: We first show that, for any $\epsilon>0$ and for any $j=1,\dots,J_1+J_2$,

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

For this, define the following sets:

$$\mathcal{M}_j\equiv\{m_j(\cdot,\theta)/\sigma_{P,j}(\theta):\theta\in\Theta,P\in\mathcal{P}\} \tag{H.199}$$
$$\mathcal{S}_j\equiv\{(m_j(\cdot,\theta)/\sigma_{P,j}(\theta))^2:\theta\in\Theta,P\in\mathcal{P}\}. \tag{H.200}$$

By Assumptions E.1-(a), E.1 (iv), E.5 (i), (iii), and arguing as in the proof of Lemma D.2.2 (and D.2.1) in Bugni, Canay, and Shi (2015b), 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 $\epsilon > 0$,

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma^2_{P,j}(\theta)} - \frac{E_P[m_j(X,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big| \leq \epsilon\Big) \to 1 \tag{H.201}$$

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta) - E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big| \leq \epsilon\Big) \to 1. \tag{H.202}$$

Note that, by Assumption E.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 (H.202), for any $\epsilon > 0$, it follows that

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta)^2 - E_P[m_j(X,\theta)]^2}{\sigma^2_{P,j}(\theta)}\Big| \leq \epsilon\Big) \to 1. \tag{H.203}$$

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

$$\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)} - 1\Big| \leq \eta \;\Rightarrow \Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - 1\Big| \leq \epsilon. \tag{H.204}$$

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

$$\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)} - 1\Big| \leq \Big|\frac{n^{-1}\sum_{i=1}^n m(X_i,\theta)^2 - E[m_j(X_i,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big| + \Big|\frac{\bar{m}_{n,j}(\theta)^2 - E[m_j(X_i,\theta)]^2}{\sigma^2_{P,j}(\theta)}\Big|. \tag{H.205}$$

By (H.204)-(H.205), bounding each of the terms on the right hand side of (H.205) by $\eta/2$ implies $|\hat{\sigma}_{n,j}(\theta)/\sigma_{P,j}(\theta) - 1| \leq \epsilon$. This, together with (H.201) and (H.203), ensures that, for any $\epsilon > 0$, (H.198) 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 (2015b) to conclude that

$$\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}\Big|\frac{\sigma_{P,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} - 1\Big| \leq \epsilon\Big) \to 1. \tag{H.206}$$

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

$$\begin{aligned}
1 &= \lim_{n\to\infty}\inf_{P\in\mathcal{P}} P\Big(\sup_{m\geq n}\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \leq \epsilon\Big) \\
&\leq \inf_{P\in\mathcal{P}}\lim_{n\to\infty} P\Big(\bigcap_{m\geq n}\big\{\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \leq \epsilon\big\}\Big) \\
&= \inf_{P\in\mathcal{P}} P\Big(\lim_{n\to\infty}\bigcap_{m\geq n}\{\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \leq \epsilon\}\Big) \\
&= \inf_{P\in\mathcal{P}} P\Big(\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \leq \epsilon,\ \text{for almost all } n\Big),
\end{aligned} \tag{H.207}$$

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 (H.14)-(H.15).

Note that

$$\begin{aligned}
&\sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}\Big|\kappa_n^{-1}\frac{\sqrt{n}\bar{m}_{n,j}(\theta_n')}{\hat{\sigma}_{n,j}(\theta_n')}-\kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\Big| \qquad \text{(H.208)}\\
&\le \sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}\Big|\kappa_n^{-1}\frac{\sqrt{n}(\bar{m}_{n,j}(\theta_n')-E_{P_n}[m_j(X_i,\theta_n')])}{\sigma_{n,j}(\theta_n')}(1+\eta_{n,j}(\theta_n'))\\
&\quad+\kappa_n^{-1}\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\Big|\\
&\le \sup_{\theta_n'\in\theta_n+\rho/\sqrt{n}B^d}|\kappa_n^{-1}\mathbb{G}_n(\theta_n')(1+\eta_{n,j}(\theta_n'))|+\Big|\frac{\sqrt{n}E_{P_n}[m_j(X_i,\theta_n')]}{\kappa_n\sigma_{P_n,j}(\theta_n')}\eta_{n,j}(\theta_n')\Big|=o_{\mathcal{P}}(1),
\end{aligned}$$

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 (2015b), 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,

$$\hat{\mu}_{n,j}(\theta_n)\overset{P_n}{\to}1-\min\Big\{\max(0,\frac{\pi_{1,j}}{\pi_{1,j+R_1}+\pi_{1,j}}),1\Big\}, \qquad \text{(H.209)}$$

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

$$\hat{\mu}_{n,j}(\theta_n)\overset{P_n}{\to}1. \qquad \text{(H.210)}$$

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

$$\hat{\mu}_{n,j+R_1}(\theta_n)\overset{P_n}{\to}1. \qquad \text{(H.211)}$$

Now, one may write

$$\begin{aligned}
\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}-1&=\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)}\Big(\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}-1\Big)+\Big(\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)}-1\Big) \qquad \text{(H.212)}\\
&=O_{P_n}(1)\Big(\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}-1\Big)+o_{P_n}(1),
\end{aligned}$$

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,

$$\begin{aligned}
\hat{\sigma}^M_{n,j}(\theta_n)&=\hat{\mu}_{n,j}(\theta_n)\hat{\sigma}_{n,j}(\theta_n)+(1-\hat{\mu}_{n,j}(\theta_n))\hat{\sigma}_{n,j+R_1}(\theta_n) \qquad \text{(H.213)}\\
&=(1+o_{P_n}(1))\hat{\sigma}_{n,j}(\theta_n)+(1-\hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)), \qquad \text{(H.214)}
\end{aligned}$$

where the second equality follows from (H.210) and the fact that

$$\text{(H.215)}\quad \hat{\sigma}_{n,j+R_1}(\theta_n) = \Big(\hat{\sigma}^2_{n,j}(\theta_n) + 2\widehat{Cov}_n(m_j(X_i,\theta), t_j(X_i,\theta)) + \widehat{Var}_n(t_j(X_i,\theta))\Big)^{1/2}$$
$$= \Big(\hat{\sigma}^2_{n,j}(\theta_n) + O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)) + O_{P_n}(1)\Big)^{1/2} = O_{P_n}(\hat{\sigma}_{n,j}(\theta_n)),$$

where the second equality follows from, $Var_{P_n}(t_j(X_i,\theta))$ being bounded by Assumption E.3-(II) and

$$\text{(H.216)}\quad \widehat{Var}_n(t_j(X_i,\theta)) = Var_{P_n}(t_j(X_i,\theta)) + o_{P_n}(1)$$
$$\text{(H.217)}\quad \widehat{Cov}_n(m_j(X_i,\theta), t_j(X_i,\theta)) \le \hat{\sigma}_{n,j}(\theta_n)\widehat{Var}_n(t_j(X_i,\theta))^{1/2},$$

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

Therefore,

$$\text{(H.218)}\quad \frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1 = \frac{\hat{\sigma}_{n,j}(\theta_n) - \hat{\sigma}^M_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}$$
$$= \frac{(1-\hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n))}{(1+o_{P_n}(1))\hat{\sigma}_{n,j}(\theta_n) + (1-\hat{\mu}_{n,j}(\theta_n))O_{P_n}(\hat{\sigma}_{n,j}(\theta_n))} = o_{P_n}(1),$$

where we used $\hat{\sigma}^{-1}_{n,j}(\theta_n) = O_{P_n}(1)$ by equation (E.3) and part (i) of the lemma. By (H.212) and (H.218), $\sigma_{P_n,j}(\theta_n)/\hat{\sigma}^M_{n,j}(\theta_n) - 1 = o_{P_n}(1)$. 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 H.9 (1),

$$\text{(H.219)}\quad \Big|\frac{\sigma_{P_n,j}(\theta_n)}{\sigma_{P_n,j+R_1}(\theta_n)} - 1\Big| = o_{P_n}(1), \quad \text{for } j = 1,\dots,R_1,$$

and therefore,

$$\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)} - 1 = \frac{\sigma_{P_n,j}(\theta_n) - \hat{\sigma}^M_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}$$
$$= \frac{[\hat{\mu}_{n,j}(\theta_n) + (1-\hat{\mu}_{n,j}(\theta_n))]\sigma_{P_n,j}(\theta_n) - [\hat{\mu}_{n,j}(\theta_n)\hat{\sigma}_{n,j}(\theta_n) + (1-\hat{\mu}_{n,j}(\theta_n))\hat{\sigma}_{n,j+R_1}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)}$$
$$\text{(H.220)}\quad = \frac{\hat{\mu}_{n,j}(\theta_n)[\sigma_{P_n,j}(\theta_n) - \hat{\sigma}_{n,j}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)} + \frac{(1-\hat{\mu}_{n,j}(\theta_n))[\sigma_{P_n,j+R_1}(\theta_n) - \hat{\sigma}_{n,j+R_1}(\theta_n) + o_{P_n}(1)]}{\hat{\sigma}^M_{n,j}(\theta_n)},$$

where the second equality follows from the definition of $\hat{\sigma}^M_{n,j}(\theta_n)$, and the third equality follows from (H.219) and $\sigma_{P_n,j+R_1}$ bounded away from 0 due to (E.3). Note that

$$\text{(H.221)}\quad \frac{\hat{\mu}_{n,j}(\theta_n)[\sigma_{P_n,j}(\theta_n) - \hat{\sigma}_{n,j}(\theta_n)]}{\hat{\sigma}^M_{n,j}(\theta_n)} = \hat{\mu}_{n,j}(\theta_n)\frac{\hat{\sigma}_{n,j}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}\Big(\frac{\sigma_{P_n,j}(\theta_n)}{\hat{\sigma}_{n,j}(\theta_n)} - 1\Big) = o_{P_n}(1),$$

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

$$\begin{aligned}(\text{H.222})\quad &\frac{(1-\hat{\mu}_{n,j}(\theta_n))[\sigma_{P_n,j+R_1}(\theta_n)-\hat{\sigma}_{n,j+R_1}(\theta_n)+o_{P_n}(1)]}{\hat{\sigma}^M_{n,j}(\theta_n)}\\ &\qquad =(1-\hat{\mu}_{n,j}(\theta_n))\frac{\hat{\sigma}_{n,j+R_1}(\theta_n)}{\hat{\sigma}^M_{n,j}(\theta_n)}\Big(\frac{\sigma_{P_n,j+R_1}(\theta_n)}{\hat{\sigma}_{n,j+R_1}(\theta_n)}-1+o_{P_n}(1)\Big)=o_{P_n}(1).\end{aligned}$$

By (H.220)-(H.222), 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. *Q.E.D.*

## *H. Lemmas Used to Prove Theorem D.1*

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

$$\begin{aligned}\mathfrak{G}^b_{n,j}(\theta) &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(m_j(X^b_i,\theta)-\bar{m}_n(\theta)\right)/\sigma_{P,j}(\theta)\\ (\text{H.223})\qquad &= \frac{1}{\sqrt{n}}\sum_{i=1}^n(M_{n,i}-1)m_j(X_i,\theta)/\sigma_{P,j}(\theta),\end{aligned}$$

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 H.3 for details on the construction of the bootstrapped empirical process).

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

$$(\text{H.224})\qquad \sup_Q N(\epsilon\|F\|_{L^2_Q},\mathcal{M}_P,L^2_Q)\le K\epsilon^{-v},\ \ 0<\epsilon<1,$$

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

$$(\text{H.225})\qquad \|(\theta_1,\tilde{\theta}_1)-(\theta_2,\tilde{\theta}_2)\|\le\delta\quad\Rightarrow\quad\sup_{P\in\mathcal{P}}\|Q_P(\theta_1,\tilde{\theta}_1)-Q_P(\theta_2,\tilde{\theta}_2)\|\le M\delta^\gamma.$$

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

$$(\text{H.226})\qquad \sup_{P\in\mathcal{P}}P\left(\sup_{\|\theta-\theta'\|\le\delta_n}\|\mathbb{G}_n(\theta)-\mathbb{G}_n(\theta')\|>\epsilon_n\right)=o(1).$$

*Further,*

$$(\text{H.227})\qquad \lim_{n\to\infty}P^*_n\left(\sup_{\|\theta-\theta'\|\le\delta_n}\|\mathfrak{G}^b_n(\theta)-\mathfrak{G}^b_n(\theta')\|>\epsilon_n|\{X_i\}_{i=1}^\infty\right)=0.$$

*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 $P$:

$$P\left(\sup_{\|\theta-\theta'\|\le\delta_n}\max_{j=1,\dots,J}|\mathbb{G}_{n,j}(\theta)-\mathbb{G}_{n,j}(\theta')|>\epsilon_n\right)=o(1). \tag{H.228}$$

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 (2015b), one has

$$\|\theta-\theta'\|\le\delta_n \quad\Rightarrow\quad \varrho_P(\theta,\theta')\le\tilde{\delta}_n, \tag{H.229}$$

where $\tilde{\delta}_n=O(\delta_n^{\gamma})$ by assumption. Define

$$\mathcal{M}_{P,\tilde{\delta}_n}=\{\sigma_{P,j}(\theta)^{-1}m_j(\cdot,\theta)-\sigma_{P,j}(\theta')^{-1}m_j(\cdot,\theta')|\theta,\theta'\in\Theta,\varrho_P(\theta,\tilde{\theta})<\tilde{\delta}_n,j=1,\dots,J\}. \tag{H.230}$$

Define $Z_n(\tilde{\delta}_n)\equiv\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}|\sqrt{n}(\mathbb{P}_n-P)f|$. Then, by (H.229), one has

$$P\left(\sup_{\|\theta-\theta'\|\le\delta_n}\max_{j=1,\dots,J}|\mathbb{G}_{n,j}(\theta)-\mathbb{G}_{n,j}(\theta'))|>\epsilon_n)\le P(Z_n(\tilde{\delta}_n)>\epsilon_n\right). \tag{H.231}$$

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

$$P(Z_n(\tilde{\delta}_n)>\epsilon_n)\le\frac{2}{\epsilon_n}E_{P\times P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\Big|\right], \tag{H.232}$$

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

$$\begin{aligned}E_{P^W}\left[\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\Big|\right]&\le K\int_0^{\tilde{\delta}_n}\sqrt{\ln N(\epsilon,\mathcal{M}_{P,\tilde{\delta}_n},L^2_{\hat{P}_n})}d\epsilon\\&\le K\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}\sup_Q\sqrt{\ln N(\epsilon\|F\|_{L^2_Q},\mathcal{M}_P,L^2_Q)}d\epsilon\\&\le K'\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}\sqrt{-v\ln\epsilon}d\epsilon,\end{aligned} \tag{H.233}$$

for some $K'>0$, where the last inequality follows from (H.224). Note that $\sqrt{-\ln\epsilon}\le-\ln\epsilon$ for $\epsilon\le\tilde{\delta}_n/\|F\|_{L^2_Q}$ with $n$ sufficiently large. Hence,

$$E_{P^W}\left[\sup_{f\in\mathcal{M}_{\tilde{\delta}_n}}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i)\Big|\right]\le K'v^{1/2}\int_0^{\tilde{\delta}_n/\|F\|_{L^2_Q}}(-\ln\epsilon)d\epsilon=K'v^{1/2}(\tilde{\delta}_n-\tilde{\delta}_n\ln(\tilde{\delta}_n)). \tag{H.234}$$

By (H.232) and taking expectations with respect to $P$ in (H.234), it follows that

$$P(Z_n(\tilde{\delta}_n) > \epsilon_n) \le 2K' v^{1/2}(\tilde{\delta}_n - \tilde{\delta}_n \ln(\tilde{\delta}_n))/\epsilon_n = O(\delta_n^{\gamma}/\epsilon_n) + O(|\delta_n^{\gamma} \ln(\delta_n)|/\epsilon_n) = o(1), \tag{H.235}$$

where the last equality follows from the rate condition on $\epsilon_n$. By (H.231) and (H.235), 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 (H.229), one has

$$P_n^*\left(\sup_{\|\theta-\theta'\|\le\delta_n} \max_{j=1,\dots,J} |\mathfrak{G}_{n,j}^b(\theta) - \mathfrak{G}_{n,j}^b(\theta')| > \epsilon_n \Big| \{X_i\}_{i=1}^{\infty}\right) \le P_n^*\big(Z_n^*(\tilde{\delta}_n) > \epsilon_n \big| \{X_i\}_{i=1}^{\infty}\big). \tag{H.236}$$

By Markov's inequality and Lemma 2.3.1 (symmetrization lemma) in van der Vaart and Wellner (2000), one has

$$P_n^*\big(Z_n^*(\tilde{\delta}_n) > \epsilon_n \big| \{X_i\}_{i=1}^{\infty}\big) \le \frac{2}{\epsilon_n} E_{P_n^*\times P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i^b)\Big| \Big| \{X_i\}_{i=1}^{\infty}\right] \tag{H.237}$$

$$= \frac{2}{\epsilon_n} E_{P_n^*}\left[E_{P^W}\left[\sup_{f\in\mathcal{M}_{P,\tilde{\delta}_n}} \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i f(X_i^b)\Big| \Big| \{X_i^b\}, \{X_i\}_{i=1}^{\infty}\right] \Big| \{X_i\}_{i=1}^{\infty}\right], \tag{H.238}$$

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 (H.232)-(H.235). 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. *Q.E.D.*

LEMMA H.12 *Suppose Assumptions E.1, E.2, and E.5 hold. Let $\mathcal{S}_P \equiv \{f : \mathcal{X} \to \mathbb{R} : f(\cdot) = \sigma_{P,j}(\theta)^{-2} m_j^2(\cdot,\theta), \theta \in \Theta, j = 1,\dots,J\}$ and let $F$ be its envelope. (i) If $\mathcal{S}_P$ is Donsker and pre-Gaussian uniformly in $P \in \mathcal{P}$, then*

$$\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* = O_{\mathcal{P}}(1/\sqrt{n}); \tag{H.239}$$

*(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 $\theta, \theta' \in \Theta$, $x \in \mathcal{X}$, $j = 1,\dots,J$, and $P \in \mathcal{P}$, then, for any $\eta > 0$, there exists a constant $C > 0$ such that*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\Big(\max_{j=1,\dots,J} \sup_{\|\theta-\theta'\|<\delta} |\eta_{n,j}(\theta) - \eta_{n,j}(\theta')| > C\delta\Big) < \eta. \tag{H.240}$$

PROOF: We show the claim by first showing that, for any $\delta > 0$, there exist $M > 0$ and $N \in \mathbb{N}$ such that

$$\inf_{P\in\mathcal{P}} P^{\infty}\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - 1\Big| \le M/\sqrt{n}\Big) \ge 1 - \delta, \ \forall n \ge N. \tag{H.241}$$

By Assumptions E.1 (iv), E.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 $\epsilon > 0$,

$$\Big|P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma^2_{P,j}(\theta)} - \frac{E_P[m_j(X,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big| \le C_1\Big) - \Pr(\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)| \le C_1)\Big| \le \epsilon \tag{H.242}$$

$$\Big|P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta) - E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big| \le C_2\Big) - \Pr(\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)| \le C_2)\Big| \le \epsilon. \tag{H.243}$$

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 H.10 (i), bounding each term of the right hand side of (H.205) by $C_1/\sqrt{n}$ and $C_2/\sqrt{n}$ implies that $\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)} - 1\Big| \le C/\sqrt{n}$ for some constant $C > 0$. Now choose $C_1 > 0$ and $C_2 > 0$ so that

$$\Pr(\sup_{\theta\in\Theta}|\mathbb{H}_{P,j}(\theta)| \le C_1) > 1 - \delta/3 \quad\text{and}\quad \Pr(\sup_{\theta\in\Theta}|\mathbb{G}_{P,j}(\theta)| \le C_2) > 1 - \delta/3 \tag{H.244}$$

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 $n$,

$$\begin{aligned} 1 - \delta &\le P\Big(\sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{n^{-1}\sum_{i=1}^n m_j(X_i,\theta)^2}{\sigma^2_{P,j}(\theta)} - \frac{E_P[m_j(X,\theta)^2]}{\sigma^2_{P,j}(\theta)}\Big| \le C_1, \\ &\qquad \sqrt{n}\sup_{\theta\in\Theta}\Big|\frac{\bar{m}_{n,j}(\theta) - E_P[m_j(X,\theta)]}{\sigma_{P,j}(\theta)}\Big| \le C_2\Big) & \text{(H.245)} \\ &\le P\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)} - 1\Big| \le C/\sqrt{n}\Big), & \text{(H.246)} \end{aligned}$$

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 (H.246), for sufficiently large $n$,

$$1 - \delta \le P\Big(\sup_{\theta\in\Theta}\Big|\frac{\hat{\sigma}_{n,j}(\theta)}{\sigma_{P,j}(\theta)} - 1\Big| \le C/\sqrt{n}\Big), \tag{H.247}$$

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

$$P\Big(\sup_{\theta\in\Theta}|\eta_{n,j}(\theta)| \le C'/\sqrt{n}\Big) \ge 1 - \delta, \tag{H.248}$$

uniformly in $P \in \mathcal{P}$ for all $n$ sufficiently large. This establishes the first claim of the lemma.

(ii) First, consider

$$\frac{\hat{\sigma}^2_{n,j}(\theta)}{\sigma^2_{P,j}(\theta)} = n^{-1}\sum_{i=1}^n\left(\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2 - \left(n^{-1}\sum_{i=1}^n\frac{m(X_i,\theta)}{\sigma_{P,j}(\theta)}\right)^2. \tag{H.249}$$

We claim that this function is Lipschitz with probability approaching 1. To see this, note that, for any $\theta, \theta' \in \Theta$,

$$
\begin{aligned}
&\left| n^{-1} \sum_{i=1}^{n} \left( \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right)^2 - n^{-1} \sum_{i=1}^{n} \left( \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right)^2 \right| \\
&= \left| n^{-1} \sum_{i=1}^{n} \left( \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} + \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right) \left( \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} - \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right) \right| \\
&\leq n^{-1} \sum_{i=1}^{n} 2 \sup_{\theta \in \Theta} \left| \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right| \bar{M}(X_i) \|\theta - \theta'\|. \qquad \text{(H.250)}
\end{aligned}
$$

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,

$$
P(B_n > C) \leq \frac{E[B_n]}{C} \leq \frac{2 E_P \left[ \sup_{\theta \in \Theta} \left| \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right|^2 \right]^{1/2} E_P \left[ \bar{M}(X_i)^2 \right]^{1/2}}{C} \leq \frac{2M}{C}, \qquad \text{(H.251)}
$$

where the third inequality is due to Assumptions E.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 $\theta, \theta' \in \Theta$,

$$
\begin{aligned}
&\left| \left( n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right)^2 - \left( n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right)^2 \right| \\
&= \left| n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} + n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right| \left| n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} - n^{-1} \sum_{i=1}^{n} \frac{m(X_i, \theta')}{\sigma_{P,j}(\theta')} \right| \\
&\leq n^{-1} \sum_{i=1}^{n} 2 \sup_{\theta \in \Theta} \left| \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right| n^{-1} \sum_{i=1}^{n} \bar{M}(X_i) \|\theta - \theta'\|. \qquad \text{(H.252)}
\end{aligned}
$$

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,

$$
\begin{aligned}
P(\tilde{B}_n > C) &\leq \frac{E[\tilde{B}_n]}{C} \leq \frac{2 E_P \left[ \left( n^{-1} \sum \sup_{\theta \in \Theta} \left| \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right| \right)^2 \right]^{1/2} E_P \left[ \left( n^{-1} \sum \bar{M}(X_i) \right)^2 \right]^{1/2}}{C} \\
&\leq \frac{2 E_P \left[ \sup_{\theta \in \Theta} \left| \frac{m(X_i, \theta)}{\sigma_{P,j}(\theta)} \right|^2 \right]^{1/2} E_P [\bar{M}(X_i)^2]^{1/2}}{C} \leq \frac{2M}{C}, \qquad \text{(H.253)}
\end{aligned}
$$

where the last inequality is due to Assumptions E.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 $|g(y) - g(y')| \leq \frac{1}{2} \sup_{\bar{y} \in (1-\epsilon, 1+\epsilon)} |\bar{y}|^{-3/2} |y - y'|$ on $(1 - \epsilon, 1 + \epsilon)$. As shown in (H.247), $\hat{\sigma}_{n,j}^2(\theta)/\sigma_{P,j}^2(\theta)$ converges to 1 in probability, and $g$ is locally Lipschitz on a neighborhood of 1. Combining this with (H.249)-(H.253) yields the desired result. *Q.E.D.*

LEMMA H.13 *Suppose Assumption E.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* $\theta, \theta' \in \Theta$, $x \in \mathcal{X}$, $j = 1, \dots, J$, *and* $P \in \mathcal{P}$. *Then,*

$$\sup_{P\in\mathcal{P}} \|Q_P(\theta_1, \tilde{\theta}_1) - Q_P(\theta_2, \tilde{\theta}_2)\| \le M\|(\theta_1, \tilde{\theta}_1) - (\theta_2, \tilde{\theta}_2)\|, \tag{H.254}$$

*for some* $M > 0$ *and for all* $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$.

PROOF: Recall that

$$[Q_P(\theta_1, \tilde{\theta}_1)]_{j,k} = E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big]. \tag{H.255}$$

For any $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$,

$$\begin{aligned}
&\Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le \Big|E_P\Big[\Big(\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)} - \frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big)\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \qquad \text{(H.256)}\\
&\qquad + \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big(\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)} - \frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big)\Big]\Big| \\
&\le E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_k(X_i,\theta)}{\sigma_{P,k}(\theta)}\Big|\bar{M}(X_i)\Big]\|\theta_1 - \theta_2\| + E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_j(X_i,\theta)}{\sigma_{P,j}(\theta)}\Big|\bar{M}(X_i)\Big]\|\tilde{\theta}_1 - \tilde{\theta}_2\| \\
&\le M(\|\theta_1 - \theta_2\| + \|\tilde{\theta}_1 - \tilde{\theta}_2\|), \qquad \text{(H.257)}
\end{aligned}$$

where the last inequality is due to the Cauchy-Schwarz inequality, Assumption E.1 (iv), and the assumption on $\bar{M}$.

Similarly, for any $\theta_1, \tilde{\theta}_1, \theta_2, \tilde{\theta}_2 \in \Theta$,

$$\begin{aligned}
&\Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)}\Big] - E_P\Big[\frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big]E_P\Big[\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)} - \frac{m_j(X_i,\theta_2)}{\sigma_{P,j}(\theta_2)}\Big]\Big|\Big|E_P\Big[\frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\qquad + \Big|E_P\Big[\frac{m_j(X_i,\theta_1)}{\sigma_{P,j}(\theta_1)}\Big]\Big|\Big|E_P\Big[\frac{m_k(X_i,\tilde{\theta}_1)}{\sigma_{P,k}(\tilde{\theta}_1)} - \frac{m_k(X_i,\tilde{\theta}_2)}{\sigma_{P,k}(\tilde{\theta}_2)}\Big]\Big| \\
&\le E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_k(X_i,\theta)}{\sigma_{P,k}(\theta)}\Big|\Big]E_P[\bar{M}(X_i)]\|\theta_1 - \theta_2\| + E_P\Big[\sup_{\theta\in\Theta}\Big|\frac{m_j(X_i,\theta)}{\sigma_{P,j}(\theta)}\Big|\Big]E_P[\bar{M}(X_i)]\|\tilde{\theta}_1 - \tilde{\theta}_2\| \\
&\le M(\|\theta_1 - \theta_2\| + \|\tilde{\theta}_1 - \tilde{\theta}_2\|), \qquad \text{(H.258)}
\end{aligned}$$

where the last inequality is due to the Cauchy-Schwarz inequality, Assumption E.1 (iv), and the assumption on $\bar{M}$. The conclusion of the lemma then follows from (H.255)-(H.258). *Q.E.D.*

## *H. Almost Sure Representation Lemma and Related Results*

In this appendix, we provide details on the almost sure representation used in Lemmas H.3, H.4, H.6, and H.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 $X^\infty = (X_1, X_2, \dots)$ 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 $X^\infty$ as the coordinate projection on the first $\infty$ 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 $(n; 1/n, \dots, 1/n)$ and is independent of $X^\infty$. 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 $\Theta$. We simply write $\ell^\infty(\Theta)$ when $J = 1$.

Using the multinomial weight, we rewrite the empirical bootstrap process as

$$\mathbb{G}^b_{n,j}(\cdot) = g_j(X^\infty, M_n) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (M_{n,i} - 1) m_j(X_i, \cdot)/\hat{\sigma}_{n,j}(\cdot), \; j = 1, \dots, J, \tag{H.259}$$

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 $(X^\infty, M_n)$ 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 $g = (g_1, \dots, g_J)'$. 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 $X^\infty$ is

$$E_M[f(\mathbb{G}^b_n)|X^\infty = x^\infty] = \int f \circ g(x^\infty, m_n) dQ(m_n), \tag{H.260}$$

where, with a slight abuse of notation, we use $Q$ for the induced law of $M_n$.

Let $\mathcal{F}$ be the function space $\{f(\cdot) = (m_1(\cdot, \theta)/\sigma_{P,1}(\theta), \dots, m_J(\cdot, \theta)/\sigma_{P,J}(\theta)), \theta \in \Theta, P \in \mathcal{P}\}$. For each $j$, define a bootstrapped empirical process standardized by $\sigma_{P,j}$ as follows:

$$\begin{aligned}
\mathfrak{G}^b_{n,j}(\theta) &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(m_j(X^b_i, \theta) - \bar{m}_n(\theta)\right)/\sigma_{P,j}(\theta) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (M_{n,i} - 1) m_j(X_i, \theta)/\sigma_{P,j}(\theta). \qquad \text{(H.261)}
\end{aligned}$$

The following result was shown in the proof of Lemma D.2.8 in Bugni, Canay, and Shi (2015b), 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 H.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,*

$$\sup_{h \in BL_1} |E_M[h(\mathfrak{G}^b_n)|X^\infty] - E[h(\mathbb{G}_P)]| \overset{as*}{\to} 0, \tag{H.262}$$

*uniformly in $P \in \mathcal{P}$.*

The result above gives uniform consistency of the standardized bootstrap process $\mathfrak{G}_n^b$. We now extend this to the studentized bootstrap process $\mathbb{G}_n^b$.

LEMMA H.15 *Suppose Assumptions E.1, E.2, and E.5 hold. Then,*

$$\sup_{h\in BL_1} |E_M[h(\mathbb{G}_n^b)|X^\infty] - E[h(\mathbb{G}_P)]| \overset{as*}{\to} 0, \tag{H.263}$$

*uniformly in* $P \in \mathcal{P}$.

PROOF: By Assumptions E.1 (iv) and E.5, Assumptions A.1-A.4 in Bugni, Canay, and Shi (2015b) 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 E.1 (iv) again, $\sup_{P\in\mathcal{P}} P^*\|f - Pf\|_{\mathcal{F}} < \infty$. Hence, by Lemma H.14,

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\sup_{h\in BL_1} |E_M[h(\mathfrak{G}_n^b)|X^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big) = 1. \tag{H.264}$$

For later use, we define the following set of sample paths, which has probability 1 uniformly in $P \in \mathcal{P}$.

$$A \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{h\in BL_1} |E_M[h(\mathfrak{G}_n^b)|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big\}. \tag{H.265}$$

Note that $\mathbb{G}_{n,j}^b$ and $\mathfrak{G}_{n,j}^b$ are related to each other by the following relationship:

$$\mathbb{G}_{n,j}^b(\theta) - \mathfrak{G}_{n,j}^b(\theta) = \mathfrak{G}_{n,j}^b(\theta)\left(\frac{\sigma_{P,j}(\theta)}{\hat\sigma_{n,j}(\theta)} - 1\right) = \mathfrak{G}_{n,j}^b(\theta)\eta_{n,j}(\theta),\ \theta \in \Theta. \tag{H.266}$$

By Assumptions E.1, E.2, and E.5, Lemma H.10 applies. Hence,

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* \to 0\Big) = 1. \tag{H.267}$$

Define the following set of sample paths:

$$B \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{\theta\in\Theta} |\eta_{n,j}(\theta)|^* \to 0, \forall j = 1, \dots, J\Big\}. \tag{H.268}$$

For any $x^\infty \in A \cap B$, it then follows that

$$\sup_{h\in BL_1} \big|E_M[h(\mathbb{G}_n^b)|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]\big|^* \to 0, \tag{H.269}$$

due to (H.264) and (H.266), $h$ being Lipschitz, $\mathfrak{G}_{n,j}^b$ being bounded (given $x^\infty$), 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 (H.264), (H.267), and De Morgan's law. This establishes the conclusion of the lemma. *Q.E.D.*

The following lemma shows that, for almost all sample path $x^\infty$, one can find an almost sure representation of the bootstrapped empirical process that is convergent.

LEMMA H.16 *Suppose Assumptions E.1, E.2, and E.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*

$$\int h \circ g(x^\infty, m_n) dQ(m_n) = \int h(\tilde{G}_{n,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1 \tag{H.270}$$

$$\int h(\mathbb{G}_P(\omega)) dP(\omega) = \int h(\tilde{G}_{P,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1, \tag{H.271}$$

*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*

$$\inf_{P \in \mathcal{P}} P^\infty\Big(\big\{x^\infty \in \mathcal{X}^\infty : \tilde{G}_{n,x^\infty} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{G}_{P,x^\infty}\big\}\Big) = 1. \tag{H.272}$$

PROOF: Define the following set of sample paths:

$$C \equiv \Big\{x^\infty \in \mathcal{X}^\infty : \sup_{h \in BL_1} |E_M[h(\mathbb{G}^b_{n,j})|X^\infty = x^\infty] - E[h(\mathbb{G}_P)]|^* \to 0\Big\}. \tag{H.273}$$

By Lemma H.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 (H.259). 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 H.15,

$$\sup_{h \in BL_1} \left| \int h \circ g(x^\infty, m_n) dQ(m_n) - E[h(\mathbb{G}_P)] \right|^* \to 0. \tag{H.274}$$

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

$$\int h \circ g(x^\infty, m_n) dQ(m_n) = \int h(\tilde{G}_{n,x^\infty}(\tilde{\omega})) d\tilde{\mathbf{P}}^*(\tilde{\omega}), \ \forall h \in BL_1. \tag{H.275}$$

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

$$\tilde{G}_{n,x^\infty}(\tilde{\omega}) = g(x^\infty, M_n(\phi_n(\tilde{\omega}))), \tag{H.276}$$

for some perfect map $\phi_n : \tilde{\Omega} \to \mathcal{Z}$ (see the construction of $\phi_\alpha$ 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. (H.269), and $\inf_{P \in \mathcal{P}} P(C) = 1$, it follows that

$$\inf_{P \in \mathcal{P}} P^\infty\Big(\big\{x^\infty \in \mathcal{X}^\infty : \tilde{G}_{n,x^\infty} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{G}_{P,x^\infty}\big\}\Big) = 1. \tag{H.277}$$

This establishes the claim of the lemma. *Q.E.D.*

LEMMA H.17 *Suppose Assumptions E.1, E.2, and E.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*

$$\inf_{P\in\mathcal{P}} P^\infty\big(\sup_{h\in BL_1} |E_M[h(W_n)|X^\infty = x^\infty] - E[h(W)]|^* \to 0\big) = 1, \tag{H.278}$$

*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* $\{W^*_{n,x^\infty} \in \mathcal{W}, n \geq 1\}$ *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*

$$E_M[h(W_n)|X^\infty = x^\infty] = \int h(W^*_{n,x^\infty}(\tilde{\omega}))d\tilde{\mathbf{P}}^*(\tilde{\omega}),\ \forall h \in BL_1 \tag{H.279}$$

$$E[h(W)] = \int h(W^*_{x^\infty}(\tilde{\omega}))d\tilde{\mathbf{P}}^*(\tilde{\omega}),\ \forall h \in BL_1, \tag{H.280}$$

*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*

$$\inf_{P\in\mathcal{P}} P^\infty\Big(\big\{x^\infty \in \mathcal{X}^\infty : W^*_{n,x^\infty} \overset{\tilde{\mathbf{P}}-as*}{\to} \tilde{W}^*_{x^\infty}\big\}\Big) = 1. \tag{H.281}$$

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 H.16 and is therefore omitted. *Q.E.D.*

REMARK H.1 *When called by the Lemmas in Appendix H, Lemma H.17 is applied, for example, with* $Y_n = (vec(\hat{D}_n(\theta'_n)), \hat{\xi}_n(\theta'_n))$ *and* $Y = (vec(D), \pi_1)$.

## 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.

ARADILLAS-LOPEZ, A., AND E. TAMER (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business & Economic Statistics*, 26(3), 261–283.

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 (2015a): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259–282.

——— (2015b): "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 (2019): "Constraint Qualifications in Partial Identification," working paper.

Magnac, T., and E. Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75, 835–864.

Mammen, E. (1992): *When Does Bootstrap Work? Asymptotic Results and Simulations*. Spinger Verlag, New York, NY.

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*, 19(4), 541–564.

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.

Rasmussen, C. E., and C. K. I. Williams (2005): *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

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.