Econometrica Supplementary Material

SUPPLEMENT TO "ASYMPTOTICALLY EFFICIENT ESTIMATION OF MODELS DEFINED BY CONVEX MOMENT INEQUALITIES"

(*Econometrica*, Vol. 82, No. 1, January 2014, 387–413)

BY HIROAKI KAIDO AND ANDRES SANTOS

IN THIS SUPPLEMENTAL MATERIAL, we include all proofs of results stated in the main text, a more detailed discussion of the examples introduced in Section 2.1, and the results of our Monte Carlo study. The proof of each main result is contained in its own appendix, which also includes a discussion of the strategy of proof and the role of the auxiliary results. The contents of the Supplemental Material are organized as follows:

Appendix A: Contains the proof of Theorem 3.2 and required auxiliary results.

Appendix B: Contains the proofs of Theorems 4.1, 4.2, Corollary 4.1, and required auxiliary results.

Appendix C: Contains the proof of Theorem 4.3 and required auxiliary results.

Appendix D: Contains the proofs of Theorems 5.1, 5.2, 5.3, and 5.4.

Appendix E: Contains the proof of Theorem 3.3, and a discussion of regularity in the incomplete linear model.

Appendix F: Discusses our Assumptions in the context of Examples 2.1, 2.2, 2.3, and 2.4.

Appendix G: Reports the results of the Monte Carlo study.

For ease of reference, the following list includes notation and definitions that will be used in the Appendix:

$a \lesssim b$	$a \leq Mb$ for some constant M.
$\|\cdot\|_F$	the Frobenius norm $\|A\|_F^2 \equiv \text{trace}\{A'A\}$.
$\|\cdot\|_o$	the operator norm for linear mappings.
$\mathbf{M}$	the set of Borel probability measures on $\mathcal{X} \subseteq \mathbf{R}^{d_X}$.
$\mathbf{M}_\mu$	for some $\mu \in \mathbf{M}$, the set $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$.
$N(Q)$	a subset of $\mathbf{M}$ that contains Q in its interior.
$N(\varepsilon, \mathcal{F}, \|\cdot\|)$	covering numbers of size ε for $\mathcal{F}$ under norm $\|\cdot\|$.
$N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|)$	bracketing numbers of size ε for $\mathcal{F}$ under norm $\|\cdot\|$.
$\mathcal{S}_i$	the arguments of $\theta \mapsto F_S^{(i)}(\int m_S(x, \theta)\, dP(x))$.
$\Xi(p, Q)$	the maximizers of $\sup_{\theta \in \Theta} \langle p, \theta \rangle$ s.t. $F(\int m(x, \theta)\, dQ(x)) \leq 0$.

APPENDIX A: PROOF OF THEOREM 3.2

This appendix contains the proof of Theorem 3.2. Several of the auxiliary results are stated in more generality than needed so that they may be employed in the derivations in Theorems 4.1 and 4.3 as well.

The proof of Theorem 3.2 proceeds by verifying the conditions of Theorem 5.2.1 in Bickel, Klassen, Ritov, and Wellner (1993), which requires two

 DOI: 10.3982/ECTA10017

key ingredients: (i) characterizing the tangent space at P, which we accomplish in Theorem A.1, and (ii) showing that $Q \mapsto \nu(\cdot, \Theta_0(Q))$ is pathwise weak-differentiable at P, which we verify in Theorem A.2. Before proceeding to the formal derivation of these results, we provide an outline of the general structure of the proof.

TANGENT SPACE—Theorem A.1:
Step 1: Lemma A.16 establishes that if $\mathbf{P}$ is open relative to $\mathbf{M}_\mu$ in the τ-topology, then the tangent space must be unrestricted. Intuitively, if $\mathbf{P}$ is open and $P \in \mathbf{P}$, then all distributions Q close to P must also be in $\mathbf{P}$. Therefore, knowing that $P \in \mathbf{P}$ does not contain information that may be exploited in estimation.
Step 2: Theorem A.1 then follows from establishing that there is a neighborhood $N(P)$ of P such that all $Q \in N(P)$ satisfy: (i) Assumption 3.6(i) (shown in Corollary A.3), (ii) Assumption 3.6(ii) (by hypothesis), (iii) Assumption 3.6(iii) (established in Lemma A.2), and (iv) Assumption 3.6(iv) (demonstrated in Lemma A.8).

DIFFERENTIABILITY—Theorem A.2:
Step 1: Exploiting Lemma A.3, Lemma A.4 first shows that $\Theta_0(P)$ has nonempty interior. Corollary A.2 then extends this result to hold for all Q in a neighborhood $N(P)$ of P.
Step 2: Next, we note that since $\Theta_0(Q)$ has nonempty interior for all $Q \in N(P)$, the support function has a saddle point representation. This is shown in Lemma A.9, which also establishes that the Lagrange multipliers are unique.
Step 3: Lemma A.14 then employs the saddle point representation, the envelope theorem, and auxiliary Lemma A.10, to show that $Q \mapsto \nu(p, \Theta_0(Q))$ is pathwise weak-differentiable at P for any $p \in \mathbb{S}^{d_\theta}$.
Step 4: Finally, Theorem A.2 is shown by extending the pointwise result of Lemma A.14. The arguments exploit the continuity of Lagrange multipliers (Lemma A.12), and an auxiliary measurability result (Lemma A.13).

LEMMA A.1: *Let $f : \mathcal{X} \times \Theta \to \mathbf{R}$ be a measurable function, bounded in $(x, \theta) \in \mathcal{X} \times \Theta$ and such that $\theta \mapsto f(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$. If Assumption* 3.2 *holds and $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ is a net in $\mathbf{M}$ with $Q_\alpha \to Q$, then*

$$\limsup_{\alpha} \sup_{\theta \in \Theta} \left| \int f(x, \theta)\, dQ_\alpha(x) - \int f(x, \theta)\, dQ(x) \right| = 0.$$

PROOF: Fix $\varepsilon > 0$ and let $N_\delta(\theta) \equiv \{\tilde{\theta} \in \Theta : \|\theta - \tilde{\theta}\| < \delta\}$. By equicontinuity, for every $\theta \in \Theta$ there is a $\delta(\theta)$ with

$$\sup_{x \in \mathcal{X}, \tilde{\theta} \in N_{\delta(\theta)}(\theta)} \left| f(x, \theta) - f(x, \tilde{\theta}) \right| < \varepsilon. \tag{A.1}$$

By compactness of Θ, there then exists a finite collection $\{\theta_1, \ldots, \theta_K\}$ such that $\{N_{\delta(\theta_i)}(\theta_i)\}_{i=1}^K$ covers Θ. Hence,

$$\begin{aligned}(\text{A.2}) \qquad & \left|\int f(x,\theta)\,dQ_\alpha(x) - \int f(x,\theta)dQ(x)\right| \\ & \quad \leq 2\varepsilon + \max_{1\leq i\leq K}\left|\int f(x,\theta_i)\big(dQ_\alpha(x) - dQ(x)\big)\right|\end{aligned}$$

for any $\theta \in \Theta$. Since ε is arbitrary and $\max_{1\leq i\leq K} |\int f(x,\theta_i)(dQ_\alpha(x) - dQ(x))| \to 0$ due to f being measurable and bounded for all θ, and $Q_\alpha \to Q$ in the τ-topology, the claim of the lemma then follows from (A.2). *Q.E.D.*

LEMMA A.2: *If Assumptions* 3.2, 3.4(i)–(ii), *and* 3.5 *hold*, *then it follows that*, *for every* $P \in \mathbf{P}$, *there is a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that*, *for all* $Q \in N(P)$, $\{\int m(x,\theta)\,dQ(x)\}_{\theta\in\Theta}$ *is compact and* $\{\int m(x,\theta)\,dQ(x)\}_{\theta\in\Theta} \subset V_0$.

PROOF: First note that Assumption 3.4(i)–(ii) and the dominated convergence theorem imply that, for any $Q \in \mathbf{M}$,

$$(\text{A.3}) \qquad \lim_{\theta_1\to\theta_2} \int m(x,\theta_1)\,dQ(x) = \int m(x,\theta_2)\,dQ(x).$$

Thus, since Θ is closed by Assumption 3.2, result (A.3) implies that the set $\mathcal{R}(Q) \equiv \{\int m(x,\theta)\,dQ(x)\}_{\theta\in\Theta}$ is closed in $\mathbf{R}^{d_m}$. Moreover, $\mathcal{R}(Q)$ is also bounded by Assumption 3.4(i), and hence we conclude that $\mathcal{R}(Q)$ is compact, which establishes the first claim of the lemma. Defining $\mathcal{R}(P)^\delta \equiv \{v \in \mathbf{R}^{d_m} : \inf_{\tilde{v}\in\mathcal{R}(P)} \|v - \tilde{v}\| < \delta\}$, it then follows from V_0 being open by Assumption 3.5, $\mathcal{R}(P)$ being compact, and Assumption 3.6(iii) that $\mathcal{R}(P) \subset V_0$. Hence, there exists a $\delta_0 > 0$ such that $\mathcal{R}(P)^{\delta_0} \subset V_0$, and the second claim of the lemma then follows from Lemma A.1 implying there exists a $N(P)$ such that $\mathcal{R}(Q) \subseteq \mathcal{R}(P)^{\delta_0}$ for all $Q \in N(P)$. *Q.E.D.*

COROLLARY A.1: *Let Assumptions* 3.2, 3.4, 3.5 *hold and* $P \in \mathbf{P}$. *Then there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that* $F(\int m(x,\cdot)\,dQ(x)) : \Theta \to \mathbf{R}^{d_F}$ *is continuously differentiable for all* $Q \in N(P)$, *and in addition*,

$$\begin{aligned}& \nabla_\theta\left\{F\left(\int m(x,\theta)\,dQ(x)\right)\right\} \\ & \quad = \nabla F\left(\int m(x,\theta)\,dQ(x)\right)\int \nabla_\theta m(x,\theta)\,dQ(x).\end{aligned}$$

PROOF: By Lemma A.2, there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that $\int m(x,\theta)\,dQ(x) \in V_0$ for all $(\theta, Q) \in \Theta \times N(P)$. For any $Q \in N(P)$ and any $1 \le i \le d_F$, Assumption 3.5 then allows us to conclude that

$$\nabla_\theta \left\{ F^{(i)}\left(\int m(x,\theta)\,dQ(x) \right) \right\} = \nabla F^{(i)}\left(\int m(x,\theta)\,dQ(x) \right) \int \nabla_\theta m(x,\theta)\,dQ(x), \tag{A.4}$$

where the exchange of order of integration and differentiation is warranted by the mean value theorem, the dominated convergence theorem, and Assumption 3.4(ii). Moreover, by Assumptions 3.4(i)–(iii) and 3.5(ii), we have

$$\lim_{\theta_n \to \theta_0} \nabla F^{(i)}\left(\int m(x,\theta_n)\,dQ(x) \right) \int \nabla_\theta m(x,\theta_n)\,dQ(x) = \nabla F^{(i)}\left(\int m(x,\theta_0)\,dQ(x) \right) \int \nabla_\theta m(x,\theta_0)\,dQ(x) \tag{A.5}$$

by the dominated convergence theorem for any $\theta_n, \theta_0 \in \Theta$. The corollary then follows from (A.4) and (A.5). *Q.E.D.*

LEMMA A.3: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *It then follows that, for every* $j \in \{1, \dots, d_\theta\}$ *and every* $\theta_0 \in \Theta_0(P)$, *there exists a* $\theta_A \in \Theta_0(P)$ *satisfying* $\theta_0^{(j)} \neq \theta_A^{(j)}$.

PROOF: The proof is by contradiction. Suppose $\theta_0 \in \Theta_0(P)$ and that, for some $\bar{j} \in \{1, \dots, d_\theta\}$, we have $\theta^{(\bar{j})} = \theta_0^{(\bar{j})}$ for all $\theta \in \Theta_0(P)$. Further define $K_i \equiv \{\theta \in \Theta : F^{(i)}(\int m(x,\theta)\,dP(x)) \le 0\}$ and, for any $A \subseteq \Theta$, let

$$\Pi_{\bar{j}}\{A\} \equiv \left\{ c \in \mathbf{R} : c = \theta^{(\bar{j})} \text{ for some } \theta \in A \right\}. \tag{A.6}$$

Since Θ is convex and $F^{(i)}(\int m(x,\cdot)\,dP(x)) : \Theta \to \mathbf{R}$ is convex by Assumptions 4.2(i), 3.6(ii), and $P \in \mathbf{P}$, it follows that K_i and $\bigcap_{i \in \mathcal{A}(\theta_0,P)} K_i$ are convex. Thus, $F^{(i)}(\int m(x,\theta_0)\,dP(x)) < 0$ for all $i \in \{1, \dots, d_F\} \setminus \mathcal{A}(\theta_0, P)$ implies

$$\left\{ \theta_0^{(\bar{j})} \right\} = \Pi_{\bar{j}} \left\{ \bigcap_{i \in \mathcal{A}(\theta_0,P)} K_i \right\}, \tag{A.7}$$

or otherwise there would be a $\theta_A \in \Theta_0(P)$ with $\theta_A^{(\bar{j})} \neq \theta_0^{(\bar{j})}$. Moreover, Corollary A.1 and $P \in \mathbf{P}$ satisfying Assumption 3.6(iv) imply $\nabla_\theta\{F^{(i)}(\int m(x,$

$\theta_0)\,dP(x))\} \neq 0$ for all $i \in \mathcal{A}(\theta_0, P)$. Hence, for each $i \in \mathcal{A}(\theta_0, P)$, there is a $\theta_i \in \Theta$ with

$$F^{(i)}\left(\int m(x, \theta_i)\,dP(x)\right) < 0 \tag{A.8}$$

due to $\theta_0 \in \Theta^o$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(i). Let $\iota: \mathcal{A}(\theta_0, P) \to \{1, \ldots, \#\mathcal{A}(\theta_0, P)\}$ be a bijection and

$$k^* \equiv \inf_{1 \leq k \leq \#\mathcal{A}(\theta_0,P)} k : \left\{\Pi_{\bar{j}}\left\{\bigcap_{i:\iota(i)\leq k} K_i\right\} = \left\{\theta_0^{(\bar{j})}\right\}\right\}, \tag{A.9}$$

where we note $2 \leq k^* \leq \#\mathcal{A}(\theta_0, P)$ due to (A.7) and $\{\Pi_{\bar{j}}\{K_i\}\}^o \neq \emptyset$ for all $i \in \mathcal{A}(\theta_0, P)$ by (A.8). Next, define

$$\bar{K} \equiv \bigcap_{i:\iota(i)\leq k^*-1} K_i, \quad K_{i^*} \equiv K_{\iota^{-1}(k^*)}. \tag{A.10}$$

Since $\Pi_{\bar{j}}\{\bar{K}\}$ is not singleton valued, there exists a $\theta_A \in \bar{K}$ with $\theta_A^{(\bar{j})} \neq \theta_0^{(\bar{j})}$. It follows that if $\bar{\theta} \in \bar{K} \cap K_{i^*}$, then $\bar{\theta} \notin K_{i^*}^o$, for otherwise $c\theta_A + (1-c)\bar{\theta} \in \bar{K} \cap K_{i^*}$ for $c \in (0, 1)$ sufficiently small, contradicting (A.9). We therefore conclude that $\bar{K} \cap K_{i^*}^o = \emptyset$, and by Theorem 5.12.3 in Luenberger (1969) that there is a $p^* \in \mathbb{S}^{d_\theta}$ such that

$$\sup_{\theta \in K_{i^*}} \langle \theta, p^* \rangle \leq \inf_{\theta \in \bar{K}} \langle \theta, p^* \rangle. \tag{A.11}$$

Further note that both the infimum and supremum in (A.11) are attained at θ_0, and that since $P \in \mathbf{P}$ must satisfy Assumption 3.6(iv), that $\{\nabla_\theta\{F^{(i)}(\int m(x, \theta_0)\,dP(x))\}\}_{i\in\mathcal{A}(\theta_0,P)}$ are linearly independent by Corollary A.1. Thus, it follows from Theorem 9.4.1 in Luenberger (1969) and $\theta_0 \in \Theta^o$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(i) that

$$\begin{aligned} 0 &= p^* + \gamma_0 \nabla_\theta\left\{F^{(\iota^{-1}(k^*))}\left(\int m(x, \theta_0)\,dP(x)\right)\right\}, \\ 0 &= p^* + \sum_{k=1}^{k^*-1} \gamma_k \nabla_\theta\left\{F^{(\iota^{-1}(k))}\left(\int m(x, \theta_0)\,dP(x)\right)\right\}, \end{aligned} \tag{A.12}$$

for some scalar $\gamma_0 \neq 0$ and vector $(\gamma_1, \ldots, \gamma_{k^*-1}) \neq 0$. However, result (A.12) and Corollary A.1 contradict $P \in \mathbf{P}$ satisfying Assumption 3.6(iv) and hence the lemma follows. *Q.E.D.*

LEMMA A.4: *If Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$, *then there exists a* $\theta_0 \in \Theta$ *such that*

$$F^{(i)}\left(\int m(x, \theta_0)\, dP(x)\right) < 0 \quad \textit{for all } 1 \le i \le d_F.$$

PROOF: Let $2^{\{1,\ldots,d_F\}}$ denote the power set of $\{1, \ldots, d_F\}$ and note that $\mathcal{A}(\cdot, P) : \Theta \to 2^{\{1,\ldots,d_F\}}$. Since $\mathcal{A}(\cdot, P)$ has finite range, there exists a collection $\{\theta_j\}_{j=1}^J$ with $J < \infty$ and $\theta_j \in \Theta_0(P)$ such that, for all $\theta \in \Theta_0(P)$,

$$\mathcal{A}(\theta, P) \in \{\mathcal{A}(\theta_j, P)\}_{j=1}^J. \tag{A.13}$$

Next, select weights $\{w_j\}_{j=1}^J$ such that $w_j > 0$ and $\sum_j w_j = 1$, and define $\theta_0 \equiv \sum_j w_j \theta_j$. By convexity, we obtain

$$F^{(i)}\left(\int m(x, \theta_0)\, dP(x)\right) \le \sum_{j=1}^J w_j F^{(i)}\left(\int m(x, \theta_j)\, dP(x)\right) \tag{A.14}$$

for any $1 \le i \le d_F$, which implies $\theta_0 \in \Theta_0(P)$. Moreover, since $w_j > 0$ for all $1 \le j \le J$, it also follows that $F^{(i)}(\int m(x, \theta_0)\, dP(x)) = 0$ if and only if $F^{(i)}(\int m(x, \theta_j)\, dP(x)) = 0$ for all $1 \le j \le J$. Thus, by (A.13), we conclude that

$$\mathcal{A}(\theta_0, P) = \bigcap_{j=1}^J \mathcal{A}(\theta_j, P) = \bigcap_{\theta \in \Theta_0(P)} \mathcal{A}(\theta, P). \tag{A.15}$$

Next, we aim to show $\mathcal{A}(\theta_0, P) = \emptyset$, which yields the claim of the lemma. Toward this end, note that, for any $1 \le i \le d_F$, if $j \in \mathcal{S}_i$, then by Lemma A.3 there exists a $\theta_A \in \Theta_0(P)$ with $\theta_0^{(j)} \ne \theta_A^{(j)}$. Thus, by convexity of Θ and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii), we obtain that $c\theta_0 + (1 - c)\theta_A \in \Theta_0(P)$ for all $c \in (0, 1)$, and

$$F^{(i)}\left(\int m\big(x, c\theta_0 + (1 - c)\theta_A\big)\, dP(x)\right) < 0. \tag{A.16}$$

Therefore, (A.15) and (A.16) imply that $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{A}(\theta_0, P)$, or equivalently that only linear constraints can be active at θ_0. Thus, Theorem 22.2 in Rockafellar (1970) then yields that either (A.17) or (A.18) must hold:

$$F^{(i)}\left(\int m(x, \theta_L)\, dP(x)\right) < 0 \quad \text{for all } i \in \mathcal{A}(\theta_0, P) \text{ for some } \theta_L \in \mathbf{R}^{d_\theta}, \tag{A.17}$$

$$\sum_{i \in \mathcal{A}(\theta_0, P)} \gamma_i \nabla_\theta \left\{F^{(i)}\left(\int m(x, \theta_0)\, dP(x)\right)\right\} = 0 \tag{A.18}$$

$$\text{for scalars } \{\gamma_i\} \text{ with } \sup_{i \in \mathcal{A}(\theta_0, P)} \gamma_i > 0.$$

However, (A.18) is not possible due to $P \in \mathbf{P}$ satisfying Assumption 3.6(iv), and hence we conclude that (A.17) must hold. Finally, since $F^{(i)}(\int m(x, \theta_0)\, dP(x)) < 0$ for all $i \in \{1, \dots, d_F\} \setminus \mathcal{A}(\theta_0, P)$ and $\theta_0 \in \Theta^o$ due to $P \in \mathbf{P}$ satisfying Assumption 3.6(i), we obtain that, for $c \in (0, 1)$ sufficiently close to 1, $F^{(i)}(\int m(x, c\theta_0 + (1-c)\theta_L)\, dP(x)) < 0$ for all $1 \le i \le d_F$. Hence, (A.15) implies $\mathcal{A}(\theta_0, P) = \emptyset$ as desired, and the claim of the lemma follows. *Q.E.D.*

LEMMA A.5: *Let Assumptions* 3.2, 3.4(i)–(ii), 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then, there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that the mapping* $(\theta, Q) \mapsto F(\int m(x, \theta)\, dQ(x))$ *is continuous at all* $(\theta, Q) \in \Theta \times N(P)$.

PROOF: Recall that, by Lemma A.2, there is $N(P) \subseteq \mathbf{M}$ such that $\int m(x, \theta)\, dQ(x) \in V_0$ for all $(\theta, Q) \in \Theta \times N(P)$. Next, let $\{\theta_\alpha, Q_\alpha\}_{\alpha \in \mathfrak{A}}$ be a net such that $(\theta_\alpha, Q_\alpha) \to (\theta_0, Q_0) \in \Theta \times N(P)$. Since $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ is bounded by Assumption 3.4(i), and $\theta \mapsto m(x, \theta)$ is equicontinuous in x by Assumption 3.4(ii), it follows from Lemma A.1 that

$$\limsup_{\alpha} \sup_{\theta \in \Theta} \left\| F\left(\int m(x, \theta)\, dQ_\alpha(x)\right) - F\left(\int m(x, \theta)\, dQ_0(x)\right) \right\| = 0, \tag{A.19}$$

due to F being uniformly continuous on V_0 by Assumption 3.5(ii). Moreover, since $\int m(x, \theta_0)\, dQ_0(x) \in V_0$, we have

$$F\left(\int m(x, \theta_\alpha)\, dQ_0(x)\right) \to F\left(\int m(x, \theta_0)\, dQ_0(x)\right) \tag{A.20}$$

by Assumption 3.4(i)–(ii) and the dominated convergence theorem. Therefore, results (A.19) and (A.20) imply that

$$F\left(\int m(x, \theta_\alpha)\, dQ_\alpha(x)\right) \to F\left(\int m(x, \theta_0)\, dQ_0(x)\right), \tag{A.21}$$

which establishes the continuity of $(\theta, Q) \to F(\int m(x, \theta)\, dQ(x))$ on $\Theta \times N(P)$ as claimed. *Q.E.D.*

COROLLARY A.2: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then, there exists a* $\theta_0 \in \Theta$ *and a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that* $F^{(i)}(\int m(x, \theta_0)\, dQ(x)) < 0$ *for all* $1 \le i \le d_F$ *and* $Q \in N(P)$.

PROOF: The claim follows immediately from Lemma A.4 implying there exists $\theta_0 \in \Theta$ such that $F^{(i)}(\int m(x, \theta_0)\, dP(x)) < 0$ for all $1 \le i \le d_F$, and Lemma A.5 implying $Q \mapsto F(\int m(x, \theta_0)\, dQ(x))$ is continuous at $Q = P$. *Q.E.D.*

LEMMA A.6: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then there is* $N(P) \subseteq \mathbf{M}$ *such that* $\Theta_0(Q) \neq \emptyset$ *is convex for all* $Q \in N(P)$, *and the correspondence* $Q \mapsto \Theta_0(Q)$ *is continuous at all* $Q \in N(P)$.

PROOF: By Θ being convex, Corollary A.2, Assumption 4.2(i), and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii), there exists a $N(P) \subseteq \mathbf{M}$ and $\theta_0 \in \Theta$ such that, for all $Q \in N(P)$ and $1 \leq i \leq d_F$, the functions $F^{(i)}(\int m(x, \cdot)\, dQ(x)) : \Theta \to \mathbf{R}$ are convex, and $F^{(i)}(\int m(x, \theta_0)\, dQ(x)) < 0$. Thus, in what follows, we let $\Theta_0(Q)$ be a convex set with nonempty interior. Moreover, by Lemma A.5, $N(P)$ may be chosen so that $(\theta, Q) \mapsto F(\int m(x, \theta)\, dQ(x))$ is continuous on $\Theta \times N(P)$.

We first establish that $Q \mapsto \Theta_0(Q)$ is lower hemicontinuous at any $Q_0 \in N(P)$. By Theorem 17.19 in Aliprantis and Border (2006), it suffices to show that, for any $\theta^* \in \Theta_0(Q_0)$ and net $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ with $Q_\alpha \to Q_0$, there exists a subnet $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ and net $\{\theta_\beta\}_{\beta \in \mathfrak{B}}$ such that $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in \mathfrak{B}$ and $\theta_\beta \to \theta^*$. If $\theta^* \in \Theta_0^o(Q_0)$, then $F^{(i)}(\int m(x, \theta^*)\, dQ_0(x)) < 0$ for all $1 \leq i \leq d_F$, and hence by Lemma A.5 and $Q_\alpha \to Q_0$, there exists α_0 such that $\theta^* \in \Theta_0(Q_\alpha)$ for all $\alpha \geq \alpha_0$. Therefore, defining $\mathfrak{B} \equiv \{\alpha \in \mathfrak{A} : \alpha \geq \alpha_0\}$, $Q_{\alpha_\beta} = Q_\beta$, and setting $\theta_\beta = \theta^*$, we obtain that $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ is a subnet with $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ and trivially satisfies $\theta_\beta \to \theta^*$. Suppose, on the other hand, that $\theta^* \in \partial\Theta_0(Q_0)$. Since $\Theta_0(Q_0)$ is convex with nonempty interior, there is a sequence $\tilde{\theta}_k$ with $\tilde{\theta}_k \to \theta^*$ and $\tilde{\theta}_k \in \Theta_0^o(Q_0)$ for all k. By Lemma A.5, there then exists a $\alpha_{0,k}$ such that $\tilde{\theta}_k \in \Theta_0(Q_\alpha)$ for all $\alpha \geq \alpha_{0,k}$. Let $\mathfrak{B} \equiv \mathfrak{A} \times \mathbb{N}$ and, for any $\beta = (\alpha, k)$, let $\alpha_\beta = \tilde{\alpha}$ for some $\tilde{\alpha} \in \mathfrak{A}$ with $\tilde{\alpha} \geq \alpha$ and $\tilde{\alpha} \geq \alpha_{0,k}$ and $\theta_\beta = \tilde{\theta}_k$. $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ is then a subnet of $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ with $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ and $\theta_\beta \to \theta^*$.

Next, we show that $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at any $Q_0 \in N(P)$. By Theorem 17.16 in Aliprantis and Border (2006), it suffices to show that any net $\{Q_\alpha, \theta_\alpha\}_{\alpha \in \mathfrak{A}}$ such that $Q_\alpha \to Q_0$ and $\theta_\alpha \in \Theta_0(Q_\alpha)$ for all $\alpha \in \mathfrak{A}$ is such that $\{\theta_\alpha\}_{\alpha \in \mathfrak{A}}$ has a limit point $\theta^* \in \Theta_0(Q_0)$. Compactness of Θ, however, implies that there exists a subnet $\{\theta_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ such that $\theta_{\alpha_\beta} \to \theta^*$ for some $\theta^* \in \Theta$. Therefore, since $\theta_{\alpha_\beta} \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in \mathfrak{B}$, we obtain

$$0 \geq F\left(\int m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)\right) \to F\left(\int m\big(x, \theta^*\big)\, dQ_0(x)\right) \tag{A.22}$$

by Lemma A.5. Thus, $\theta^* \in \Theta_0(Q_0)$ and upper hemicontinuity is established. Since, as argued, $Q \mapsto \Theta_0(Q)$ is also lower hemicontinuous, the claim of the lemma immediately follows. *Q.E.D.*

COROLLARY A.3: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then, there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that* $\emptyset \neq \Theta_0(Q) \subset \Theta^o$ *for all* $Q \in N(P)$.

PROOF: Since $\theta \mapsto F(\int m(x,\theta)\,dP(x))$ is continuous in $\theta \in \Theta$ by Lemma A.5, it follows that $\Theta_0(P)$ is closed. Hence, since $\partial\Theta$ is closed as well and $\Theta_0(P) \cap \partial\Theta = \emptyset$ due to $P \in \mathbf{P}$ satisfying Assumption 3.6(i), we must have that

$$\inf_{\theta_1 \in \Theta_0(P)} \inf_{\theta_2 \in \partial\Theta} \|\theta_1 - \theta_2\| > 0. \tag{A.23}$$

Therefore, there exists an open set U such that $\Theta_0(P) \subset U \subset \Theta^o$. Since by Lemma A.6 the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at P, there then exists a $N(P) \subseteq \mathbf{M}$ such that, for all $Q \in N(P)$, we have $\emptyset \neq \Theta_0(Q) \subset U \subset \Theta^o$; see Definition 17.2 in Aliprantis and Border (2006). *Q.E.D.*

LEMMA A.7: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and define the correspondence*

$$\Xi(p,Q) \equiv \arg\max_{\theta\in\Theta}\left\{\langle p,\theta\rangle \ s.t.\ F\left(\int m(x,\theta)\,dQ(x)\right) \le 0\right\}. \tag{A.24}$$

Then there is $N(P) \subseteq \mathbf{M}$ *with* $(p,Q) \mapsto \Xi(p,Q)$ *nonempty, compact, and upper hemicontinuous on* $\mathbb{S}^{d_\theta} \times N(P)$.

PROOF: By Lemma A.6, there exists a $N(P) \subseteq \mathbf{M}$ such that $\Theta_0(Q) \neq \emptyset$ and $Q \mapsto \Theta_0(Q)$ is continuous on $N(P)$. Since by Lemma A.5 the set $\Theta_0(Q) \subseteq \Theta$ is closed, Assumption 3.2 implies $\Theta_0(Q)$ is compact. Hence, $\Xi(p,Q)$ is well defined as the maximum is indeed attained for all $(p,Q) \in \mathbb{S}^{d_\theta} \times N(P)$. Continuity of $Q \mapsto \Theta_0(Q)$ and Theorem 17.31 in Aliprantis and Border (2006) then imply $(p,Q) \mapsto \Xi(p,Q)$ is compact valued and upper hemicontinuous. *Q.E.D.*

LEMMA A.8: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then, there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *so that* $\{\nabla F^{(i)}(\int m(x,\theta)\,dQ(x)) \int \nabla_\theta m(x,\theta)\,dQ(x)\}_{i\in\mathcal{A}(\theta,Q)}$ *are linearly independent for all* $\theta \in \Theta_0(Q)$ *and* $Q \in N(P)$.

PROOF: The proof is by contradiction. Let $\mathfrak{N}_P$ be the neighborhood system of P with direction $V \succeq W$ whenever $V \subseteq W$, which forms a directed set. If the lemma fails to hold, then, for $\mathfrak{A} = \mathfrak{N}_P$, there exists a net $\{Q_\alpha, \theta_\alpha\}_{\alpha\in\mathfrak{A}}$ such that $Q_\alpha \to P$, $\theta_\alpha \in \Theta_0(Q_\alpha)$ and the vectors $\{\nabla F^{(i)}(\int m(x,\theta_\alpha)\,dQ_\alpha(x)) \int \nabla_\theta m(x,\theta_\alpha)\,dQ_\alpha(x)\}_{i\in\mathcal{A}(\theta_\alpha,Q_\alpha)}$ are not linearly independent for all $\alpha \in \mathfrak{A}$. Since by Lemma A.6 the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous in a neighborhood of P, we may pass to a subnet $\{Q_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta\in\mathfrak{B}}$ such that $(Q_{\alpha_\beta}, \theta_{\alpha_\beta}) \to (P, \theta^*)$ with $\theta^* \in \Theta_0(P)$. Further note that for any index $i \in \mathcal{A}^c(\theta^*, P)$, Lemma A.5 implies that

$$F^{(i)}\left(\int m(x,\theta_{\alpha_\beta})\,dQ_{\alpha_\beta}(x)\right) \to F^{(i)}\left(\int m(x,\theta^*)\,dP(x)\right) < 0. \tag{A.25}$$

Therefore, there is a β_0 such that, if $\beta \geq \beta_0$, then the constraints that are inactive under (θ^*, P) are also inactive under $(\theta_{\alpha_\beta}, Q_{\alpha_\beta})$. Equivalently, for $\beta \geq \beta_0$, $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) \subseteq \mathcal{A}(\theta^*, P)$, and hence, in establishing a contradiction, it suffices to show $\{\nabla F^{(i)}(\int m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)\}_{i \in \mathcal{A}(\theta^*, P)}$ are linearly independent for some $\beta \geq \beta_0$.

Toward this end, notice that Assumption 3.4(ii)–(iii) and Lemma A.1 imply that, uniformly in $\theta \in \Theta$,

$$\int \nabla_\theta m(x, \theta)\, dQ_{\alpha_\beta}(x) \to \int \nabla_\theta m(x, \theta)\, dP(x). \tag{A.26}$$

Since $\nabla_\theta m$ is uniformly bounded and continuous in θ, the dominated convergence theorem and (A.26) yield

$$\int \nabla_\theta m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x) \to \int \nabla_\theta m\big(x, \theta^*\big)\, dP(x). \tag{A.27}$$

Similarly, since $v \mapsto \nabla F(v)$ is uniformly continuous on V_0 by Assumption 3.5(ii) and $\int m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x) \in V_0$ for β sufficiently large by Lemma A.2, Lemma A.1 applied to $\theta \mapsto m(x, \theta)$ and result (A.27) yield

$$\begin{aligned} &\nabla F\left(\int m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)\right) \int \nabla_\theta m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x) \\ &\quad \to \nabla F\left(\int m\big(x, \theta^*\big)\, dP(x)\right) \int \nabla_\theta m\big(x, \theta^*\big)\, dP(x). \end{aligned} \tag{A.28}$$

However, since $P \in \mathbf{P}$ satisfies Assumption 3.6(iv), the vectors $\{\nabla F^{(i)}(\int m(x, \theta^*)\, dP(x)) \int \nabla_\theta m(x, \theta^*)\, dP(x)\}_{i \in \mathcal{A}(\theta^*, P)}$ are linearly independent, and hence by (A.28), so must $\{\nabla F^{(i)}(\int m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x, \theta_{\alpha_\beta})\, dQ_{\alpha_\beta}(x)\}_{i \in \mathcal{A}(\theta^*, P)}$ for $\beta \geq \beta_1$ and some $\beta_1 \in \mathfrak{B}$. Thus, the contradiction is established and the claim of the lemma follows. *Q.E.D.*

LEMMA A.9: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $P \in \mathbf{P}$. *Then there is a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that, for all* $Q \in N(P)$ *and* $p \in \mathbb{S}^{d_\theta}$, *there is a unique* $\lambda(p, Q) \in \mathbf{R}^{d_F}$ *satisfying*

$$\sup_{\theta \in \Theta_0(Q)} \langle p, \theta \rangle = \sup_{\theta \in \Theta} \left\{ \langle p, \theta \rangle + \lambda(p, Q)' F\left(\int m(x, \theta)\, dQ(x)\right) \right\}. \tag{A.29}$$

PROOF: By Assumption 4.2(i), Corollary A.2, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii), there is a $N_1(P) \subseteq \mathbf{M}$ such that, for all $Q \in N_1(P)$, there is a $\theta_0 \in \Theta$ with $F^{(i)}(\int m(x, \theta_0)\, dQ(x)) < 0$ for all $1 \leq i \leq d_F$ and $F^{(i)}(\int m(x, \cdot)\, dQ(x)) :$

$\Theta \to \mathbf{R}$ is convex for all $1 \leq i \leq d_F$. Since Θ is compact and convex by Assumption 3.2, the optimization problem

$$\text{(A.30)} \quad \sup_{\theta\in\Theta}\langle p, \theta\rangle \quad \text{s.t.} \quad F\left(\int m(x, \theta)\, dQ(x)\right) \leq 0$$

satisfies the conditions of Corollary 28.2.1 in Rockafellar (1970) for all $Q \in N_1(P)$ and all $p \in \mathbb{S}^{d_\theta}$. We can therefore conclude that the equality in (A.29) holds for some $\lambda(p, Q) \in \mathbf{R}^{d_F}$.

Next, we show that there exists a $N(P) \subseteq N_1(P)$ such that $\lambda(p, Q)$ is unique for all $p \in \mathbb{S}^{d_\theta}$ and $Q \in N(P)$. To this end, note that, by Lemma A.7 and Corollary A.3, there exists a $N_2(P) \subseteq N_1(P)$ such that $\Xi(p, Q)$ as defined in (A.24) satisfies $\emptyset \neq \Xi(p, Q) \subseteq \Theta_0(Q) \subset \Theta^o$ for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N_2(Q)$. Theorem 8.3.1 in Luenberger (1969) then implies that any $\theta^* \in \Xi(p, Q)$ is also a maximizer of the dual problem, and hence, for any $\theta^* \in \Xi(p, Q)$,

$$\text{(A.31)} \quad p' + \lambda(p, Q)'\nabla F\left(\int m(x, \theta^*)\, dQ(x)\right)\int \nabla_\theta m(x, \theta^*)\, dQ(x) = 0,$$

by Corollary A.1 for all Q in some neighborhood $N_3(P) \subseteq N_2(P)$. Result (A.31) represents a linear equation in $\lambda(p, Q) \in \mathbf{R}^{d_F}$. However, by the complementary slackness conditions, $\lambda^{(i)}(p, Q) = 0$, for any $i \in \mathcal{A}^c(\theta^*, Q)$. Therefore, the linear system in equation (A.31) can be reduced to d_θ equations and $\#\mathcal{A}(\theta^*, Q)$ unknowns. Furthermore, by Lemma A.8, there is a neighborhood $N(P) \subseteq N_3(P)$ with $\{\nabla F^{(i)}(\int m(x, \theta^*)\, dQ(x)) \int \nabla_\theta m(x, \theta^*)\, dQ(x)\}_{i\in\mathcal{A}(\theta^*,Q)}$ linearly independent for all $Q \in N(P)$ and any $\theta^* \in \Theta_0(Q)$. Hence, we conclude that, for any $Q \in N(P)$, the solution to equation (A.31) in $\lambda(p, Q) \in \mathbf{R}^{d_F}$ satisfying (A.30) is unique and the claim of the lemma follows. *Q.E.D.*

LEMMA A.10: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and* $\Xi(p, Q)$ *be as in* (A.24). *Then, there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that, for each* $(Q, p) \in N(P) \times \mathbb{S}^{d_\theta}$ *and all* $1 \leq i \leq d_F$, *one of the following must hold*: (i) $\lambda^{(i)}(p, Q) = 0$, *or* (ii) $\theta_1^{(j)} = \theta_2^{(j)}$ *for all* $j \in \mathcal{S}_i$ *and all* $\theta_1, \theta_2 \in \Xi(p, Q)$.

PROOF: Recall that we refer to the *arguments* of $F_S^{(i)}(\int m_S(x, \cdot)\, dQ(x))$ as the coordinates of θ corresponding to indices in $\mathcal{S}_i$ (as in (4)). By $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) and Lemma A.7, there is a $N(P) \subseteq \mathbf{M}$ such that, for all $Q \in N(P)$ and $1 \leq i \leq d_F$, the functions $F_S^{(i)}(\int m_S(x, \cdot)\, dQ(x))$ are strictly convex in their arguments, and $\Xi(p, Q) \neq \emptyset$ for all $p \in \mathbb{S}^{d_\theta}$. To establish the lemma, we aim to show that condition (i) must hold whenever (ii) fails. To this end, suppose there exists a $1 \leq i \leq d_F$ such that $\theta_1^{(j)} \neq \theta_2^{(j)}$ for some $j \in \mathcal{S}_i$ and $\theta_1, \theta_2 \in \Xi(p, Q)$. Next, define $\theta_L = c\theta_1 + (1 - c)\theta_2$ with $c \in (0, 1)$ and note

$\theta_1^{(j)} \neq \theta_2^{(j)}$ and $j \in \mathcal{S}_i$, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) imply

$$\begin{aligned} F^{(i)}\biggl(\int m(x, \theta_L)\, dQ(x)\biggr) &\\ < cF^{(i)}\biggl(\int m(x, \theta_1)\, dQ(x)\biggr) + (1-c)F^{(i)}\biggl(\int m(x, \theta_2)\, dQ(x)\biggr) &\\ \leq 0, & \end{aligned} \tag{A.32}$$

where the second inequality follows from $\theta_1, \theta_2 \in \Theta_0(Q)$. However, since Θ is convex by Assumption 3.2, $\Theta_0(Q)$ is convex as well and hence $\theta_L \in \Theta_0(Q)$. Since $\langle p, \theta_L\rangle = c\langle p, \theta_1\rangle + (1-c)\langle p, \theta_2\rangle$, we must have $\theta_L \in \Xi(p, Q)$, and therefore (A.32) and the complementary slackness condition imply $\lambda^{(i)}(p, Q) = 0$, establishing the lemma. *Q.E.D.*

LEMMA A.11: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and* $\lambda(p, Q)$ *be as in* (A.29). *Then, there exists a* $N(P) \subseteq \mathbf{M}$ *such that* $\|\lambda(p, Q)\|$ *is uniformly bounded in* $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$.

PROOF: We establish the claim by contradiction. Let $\mathfrak{N}_P$ denote the neighborhood system of P with direction $V \succeq W$ whenever $V \subseteq W$, let $\mathbb{N}$ be the natural numbers, and note $\mathfrak{N}_P \times \mathbb{N}$ then forms a directed set. If the claim is false, then, setting $\mathfrak{A} = \mathfrak{N}_P \times \mathbb{N}$ and $\alpha = (V, k) \in \mathfrak{A}$, we may find a net $\{Q_\alpha, p_\alpha, \theta_\alpha\}_{\alpha \in \mathfrak{A}}$ such that, for all $\alpha \in \mathfrak{A}$,

$$\bigl\|\lambda(p_\alpha, Q_\alpha)\bigr\| > k, \quad Q_\alpha \in V, \quad p_\alpha \in \mathbb{S}^{d_\theta}, \quad \theta_\alpha \in \Xi(p_\alpha, Q_\alpha), \tag{A.33}$$

where $\Xi(p, Q)$ is as in (A.24). However, by: (i) $(p, Q) \mapsto \Xi(p, Q)$ being upper hemicontinuous and compact valued in a neighborhood of P, and (ii) $\mathbb{S}^{d_\theta}$ being compact, we may pass to a subnet $\{Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ such that

$$\begin{aligned} &\bigl(Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}, \bigl\|\lambda(p_{\alpha_\beta}, Q_{\alpha_\beta})\bigr\|\bigr) \to \bigl(P, p^*, \theta^*, +\infty\bigr) \\ &\quad \text{for some } \bigl(p^*, \theta^*\bigr) \in \mathbb{S}^{d_\theta} \times \Xi\bigl(p^*, P\bigr). \end{aligned} \tag{A.34}$$

Since the number of constraints is finite, there is a set of indices $\mathcal{C} \subseteq \{1, \ldots, d_F\}$ such that, for every $\beta_0 \in \mathfrak{B}$, there exists a $\beta \geq \beta_0$ with $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C}$. Letting $\mathfrak{G} \equiv \mathfrak{B}$, we may then set $\alpha_{\beta_\gamma} = \alpha_{\tilde{\beta}}$ for some $\tilde{\beta} \geq \beta$ satisfying $\mathcal{A}(\theta_{\alpha_{\tilde{\beta}}}, Q_{\alpha_{\tilde{\beta}}}) = \mathcal{C}$. In this way, we obtain a subnet which, for simplicity, we denote $\{Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}\}_{\gamma \in \mathfrak{G}}$, with

$$\begin{aligned} &\bigl(Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}, \bigl\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\bigr\|\bigr) \to \bigl(P, p^*, \theta^*, +\infty\bigr), \\ &\mathcal{A}(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) = \mathcal{C} \quad \forall \gamma \in \mathfrak{G}. \end{aligned} \tag{A.35}$$

Next, let $\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})$ and $\nabla^{\mathcal{C}} F(\int m(x, \theta_{\alpha_\gamma})\, dQ_{\alpha_\gamma}(x))$ respectively be the $\#\mathcal{C} \times 1$ vector and $\#\mathcal{C} \times d_m$ matrix that stack components of $\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})$ and $\nabla F(\int m(x, \theta_{\alpha_\gamma})\, dQ_{\alpha_\gamma}(x))$ whose indexes belong to $\mathcal{C}$. Similarly, define

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \equiv \nabla^{\mathcal{C}} F\left(\int m(x, \theta_{\alpha_\gamma})\, dQ_{\alpha_\gamma}(x)\right) \int \nabla_\theta m(x, \theta_{\alpha_\gamma})\, dQ_{\alpha_\gamma}(x). \tag{A.36}$$

By Lemma A.8, there is a γ_0 such that $M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'$ is invertible for all $\gamma \geq \gamma_0$. Therefore, since by the complementary slackness conditions $\lambda^{(i)}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = 0$ for all $i \notin \mathcal{C}$, we obtain from result (A.31) that

$$\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = -\big(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'\big)^{-1} M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) p_{\alpha_\gamma}. \tag{A.37}$$

Additionally, since $(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \to (\theta^*, P)$ as in (A.34), we obtain from result (A.28) and definition (A.36) that

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})' \to M\big(\theta^*, P\big)M\big(\theta^*, P\big)'. \tag{A.38}$$

For a symmetric matrix Σ, let $\xi(\Sigma)$ denote its smallest eigenvalue and note $\xi(M(\theta^*, P)M(\theta^*, P)') > 2\varepsilon$ for some $\varepsilon > 0$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(iv). Since eigenvalues are continuous under $\|\cdot\|_F$ by Corollary III.2.6 in Bhatia (1997), we obtain from (A.38) that there is a $\gamma_1 \geq \gamma_0 \in \mathfrak{G}$ such that, for all $\gamma \geq \gamma_1$, we have

$$\xi\big(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'\big) > \varepsilon. \tag{A.39}$$

Furthermore, since $\lambda^{(i)}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = 0$ for all $i \notin \mathcal{C}$, it follows that $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| = \|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ and hence

$$\begin{aligned}
\big\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\big\| &= \big\|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\big\| \\
&\leq \big\|\big(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'\big)^{-1}\big\|_o \\
&\quad \times \big\|M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})\big\|_F \times \|p\| \\
&\leq \xi^{-1}\big(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'\big) \times \sup_{v \in V_0} \big\|\nabla F(v)\big\|_F \\
&\quad \times \sup_{(x,\theta) \in \mathcal{X} \times \Theta} \big\|\nabla_\theta m(x, \theta)\big\|_F,
\end{aligned} \tag{A.40}$$

where the final inequality holds for all $\gamma \geq \gamma_2$ for some $\gamma_2 \in \mathfrak{G}$ with $\gamma_2 \geq \gamma_1$ by Lemma A.2. However, (A.39), (A.40), and Assumptions 3.4(ii), 3.5(ii) imply $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ is uniformly bounded for all $\gamma \geq \gamma_2$, contradicting (A.35).

Q.E.D.

LEMMA A.12: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and* $\lambda(p, Q)$ *be as in* (A.29). *Then, there exists a* $N(P) \subseteq \mathbf{M}$ *such that the function* $(p, Q) \mapsto \lambda(p, Q)$ *is continuous on* $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$.

PROOF: By Lemmas A.9 and A.11, there exists a $N_1(P) \subseteq \mathbf{M}$ such that $\lambda(p, Q)$ is well defined, unique, and uniformly bounded for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N_1(P)$. Therefore, letting $\Lambda \equiv \mathrm{cl}\{\lambda(p, Q) : (p, Q) \in \mathbb{S}^{d_\theta} \times N_1(P)\}$, it follows that Λ is compact in $\mathbf{R}^{d_F}$. By Lemma A.9 and Theorem 8.6.1 in Luenberger (1969), we then have

$$\lambda(p, Q) = \arg\min_{\lambda \geq 0} V(\lambda, p, Q) = \arg\min_{\lambda \in \Lambda} V(\lambda, p, Q), \tag{A.41}$$

$$V(\lambda, p, Q) \equiv \max_{\theta \in \Theta} \left\{ \langle p, \theta \rangle + \lambda' F\left(\int m(x, \theta)\, dQ(x) \right) \right\}.$$

Since $(\theta, Q) \mapsto F(\int m(x, \theta)\, dQ(x))$ is continuous on a neighborhood $N(P) \subseteq N_1(P)$ by Lemma A.5, compactness of Θ, and Theorem 17.31 in Aliprantis and Border (2006) imply $(\lambda, p, Q) \mapsto V(\lambda, p, Q)$ is continuous on $\Lambda \times \mathbb{S}^{d_\theta} \times N(P)$. Therefore, by (A.41), compactness of Λ and a second application of Theorem 17.31 in Aliprantis and Border (2006), it follows that $(p, Q) \mapsto \lambda(p, Q)$ is upper hemicontinuous on $\mathbb{S}^{d_\theta} \times N(P)$. However, since $(p, Q) \mapsto \lambda(p, Q)$ is a singleton valued correspondence on $\mathbb{S}^{d_\theta} \times N(P)$ by Lemma A.9, we conclude that it is, in fact, a continuous function. *Q.E.D.*

LEMMA A.13: *Let Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and* $\Xi(p, P)$ *be as in* (A.24). *Then, there exists a Borel measurable selector* $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *with* $\theta^*(p) \in \Xi(p, P)$ *for all* $p \in \mathbb{S}^{d_\theta}$.

PROOF: By Lemma A.7, $p \mapsto \Xi(p, P)$ is upper hemicontinuous in $p \in \mathbb{S}^{d_\theta}$ and hence weakly measurable; see Definition 18.1 in Aliprantis and Border (2006). Since $p \mapsto \Xi(p, P)$ is nonempty and compact valued by Lemma A.7, Theorem 18.13 in Aliprantis and Border (2006) implies there is a measurable selector $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ and the lemma follows. *Q.E.D.*

LEMMA A.14: *Let Assumptions* 3.2, 3.3, 3.4, 3.5 *hold, and* $\eta \mapsto h_\eta$ *be a curve in* $\mathbf{S}$. *Then, there is a neighborhood* $N \subseteq \mathbf{R}$ *of* 0 *such that, for all* $\eta_0 \in N$, $p \in \mathbb{S}^{d_\theta}$, $\Xi(p, P_\eta)$ *as in* (A.24), *and* $\lambda(p, P_\eta) \in \mathbf{R}^{d_F}$ *as in* (A.29),

$$\begin{aligned} &\frac{\partial}{\partial \eta} \nu\big(p, \Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0} \\ &\quad = 2\lambda(p, P_{\eta_0})' \nabla F\left(\int m\big(x, \theta^*\big) h_{\eta_0}^2(x)\, d\mu(x) \right) \\ &\quad\quad \times \int m\big(x, \theta^*\big) \dot{h}_{\eta_0}(x) h_{\eta_0}(x)\, d\mu(x) \quad \textit{for any } \theta^* \in \Xi(p, P_{\eta_0}). \end{aligned} \tag{A.42}$$

PROOF: For any $1 \leq i \leq d_m$ and $\theta \in \Theta$, first observe that by rearranging terms it follows that, for any η_0,

$$\text{(A.43)} \quad \left| \int m^{(i)}(x,\theta)\{h^2_{\eta_0}(x) - h^2_\eta(x) - 2(\eta_0 - \eta)h_{\eta_0}(x)\dot{h}_{\eta_0}(x)\}\, d\mu(x) \right|$$
$$= \left| \int m^{(i)}(x,\theta)\{(h_\eta(x) - h_{\eta_0}(x))^2 \right.$$
$$\left. + 2h_{\eta_0}(x)(h_\eta(x) - h_{\eta_0}(x) + (\eta_0 - \eta)\dot{h}_{\eta_0}(x))\}\, d\mu(x) \right|$$
$$= o(|\eta - \eta_0|),$$

where the final result holds by m being bounded by Assumption 3.4(i), Cauchy–Schwarz, $\|h_\eta - h_{\eta_0}\|^2_{L^2_\mu} = O(|\eta - \eta_0|^2)$, and $\|h_\eta - h_{\eta_0} - (\eta - \eta_0)\dot{h}_{\eta_0}\|_{L^2_\mu} = o(|\eta - \eta_0|)$ due to $\eta \mapsto h_\eta$ being Fréchet differentiable. Moreover, $\|h_\eta - h_{\eta_0}\|_{L^2_\mu} = o(1)$ implies $P_\eta \to P_{\eta_0}$ with respect to the total variation metric, and hence also with respect to the τ-topology. Thus, for η_0 in a neighborhood of zero, result (A.43), Lemma A.2, and Assumption 3.5(i)–(ii) yield

$$\text{(A.44)} \quad \frac{\partial}{\partial \eta} F\left(\int m(x,\theta)h^2_\eta(x)\, d\mu(x)\right)\Big|_{\eta=\eta_0}$$
$$= 2\nabla F\left(\int m(x,\theta)h^2_{\eta_0}(x)\, d\mu(x)\right)$$
$$\times \int m(x,\theta)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)\, d\mu(x).$$

Since $\eta \mapsto h_\eta$ is continuously Fréchet differentiable, result (A.28) implies that the derivative in (A.44) is continuous in η_0 in a neighborhood of zero. Therefore, Assumption 3.3 implying Assumption 4.2(i), Lemma A.9, and Corollary 5 in Milgrom and Segal (2002) imply $\eta \mapsto \nu(p, \Theta_0(P_\eta))$ is directionally differentiable in a neighborhood of zero, with

$$\text{(A.45)} \quad \frac{\partial}{\partial \eta_+} \nu(p, \Theta_0(P_\eta))\Big|_{\eta=\eta_0}$$
$$= \max_{\theta^* \in \Xi(p, P_{\eta_0})} 2\lambda(p, P_{\eta_0})' \nabla F\left(\int m(x,\theta^*)h^2_{\eta_0}(x)\, d\mu(x)\right)$$
$$\times \int m(x,\theta^*)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)\, d\mu(x),$$

$$\frac{\partial}{\partial\eta_-}\nu\big(p,\Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0} = \min_{\theta^*\in\Xi(p,P_{\eta_0})} 2\lambda(p,P_{\eta_0})'\nabla F\bigg(\int m\big(x,\theta^*\big)h^2_{\eta_0}(x)\,d\mu(x)\bigg) \times \int m\big(x,\theta^*\big)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)\,d\mu(x), \tag{A.46}$$

where $\frac{\partial}{\partial\eta_+}$ and $\frac{\partial}{\partial\eta_-}$ denote right and left derivatives, respectively. Note, however, that by Lemma A.10, for all $1\leq i\leq d_F$ such that $\lambda^{(i)}(p,P_{\eta_0})\neq 0$, we must have $\theta_1^{(j)}=\theta_2^{(j)}$ for all $j\in\mathcal{S}_i$ and all $\theta_1,\theta_2\in\Xi(p,P_{\eta_0})$. Therefore, since $A\theta$ trivially does not depend on η, it follows from (3), (4), and results (A.44), (A.45), and (A.46) that

$$\begin{aligned}\frac{\partial}{\partial\eta_+}\nu\big(p,\Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0} &= \max_{\theta^*\in\Xi(p,P_{\eta_0})}\sum_{i:\lambda^{(i)}(p,P_{\eta_0})\neq 0}\lambda^{(i)}(p,P_{\eta_0}) \\ &\quad\times\frac{\partial}{\partial\eta}F_S^{(i)}\bigg(\int m\big(x,\theta^*\big)h^2_\eta(x)\,d\mu(x)\bigg)\bigg|_{\eta=\eta_0} \\ &= \min_{\theta^*\in\Xi(p,P_{\eta_0})}\sum_{i:\lambda^{(i)}(p,P_{\eta_0})\neq 0}\lambda^{(i)}(p,P_{\eta_0}) \\ &\quad\times\frac{\partial}{\partial\eta}F_S^{(i)}\bigg(\int m\big(x,\theta^*\big)h^2_\eta(x)\,d\mu(x)\bigg)\bigg|_{\eta=\eta_0} \\ &= \frac{\partial}{\partial\eta_-}\nu\big(p,\Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0}.\end{aligned} \tag{A.47}$$

Thus, the claim of the lemma follows from (A.45), (A.46), and (A.47). *Q.E.D.*

LEMMA A.15: *Let Assumptions* 3.2, 3.3, 3.4, 3.5 *hold, and* $\eta\mapsto h_\eta$ *be a curve in* $\mathbf{S}$. *Then*:

(i) *there is a neighborhood* $N\subseteq\mathbf{R}$ *of* 0 *such that* $\frac{\partial}{\partial\eta}\nu(p,\Theta_0(P_\eta))|_{\eta=\eta_0}$ *is bounded in* $(p,\eta_0)\in\mathbb{S}^{d_\theta}\times N$, *and*

(ii) *the function* $(p,\eta_0)\mapsto\frac{\partial}{\partial\eta}\nu(p,\Theta_0(P_\eta))|_{\eta=\eta_0}$ *is continuous at all* $(p,\eta_0)\in\mathbb{S}^{d_\theta}\times N$.

PROOF: To establish the first claim, notice that, by Lemmas A.2, A.14, and the Cauchy–Schwarz inequality,

$$\text{(A.48)}\quad \left|\frac{\partial}{\partial\eta}\nu\big(p,\Theta_0(P_\eta)\big)\Big|_{\eta=\eta_0}\right| \le 2\big\|\lambda(p,P_{\eta_0})\big\| \times \sup_{v\in V_0}\big\|\nabla F(v)\big\|_F \times \sqrt{d_m}\sup_{(x,\theta)\in\mathcal{X}\times\Theta}\big\|m(x,\theta)\big\| \times \|\dot{h}_{\eta_0}\|_{L^2_\mu} \times \|h_{\eta_0}\|_{L^2_\mu},$$

for η_0 in a neighborhood of zero. Since $\|\dot{h}_{\eta_0}\|_{L^2_\mu}$ is continuous in η_0 due to $\eta\mapsto h_\eta$ being continuously Fréchet differentiable, it attains a finite maximum in a neighborhood of zero. Thus, $\|\dot{h}_{\eta_0}\|_{L^2_\mu}$ is uniformly bounded, and since $\|h_{\eta_0}\|_{L^2_\mu}=1$ for all η_0, Lemma A.11, Assumptions 3.4(i), 3.5(ii), and (A.48) establish the first claim of the lemma.

To establish the second claim, let $(p_n,\eta_n)\to(p_0,\eta_0)$ and select $\theta_n^*\in\Xi(p_n,P_{\eta_n})$ for $\Xi(p,Q)$ as in (A.24). Since $\|m(x,\theta)\|$ is uniformly bounded by Assumption 3.4(i), we obtain, for any $1\le i\le d_m$, that

$$\text{(A.49)}\quad \lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\int m^{(i)}(x,\theta)\big\{\dot{h}_{\eta_n}(x)h_{\eta_n}(x)-\dot{h}_{\eta_0}(x)h_{\eta_0}(x)\big\}\,d\mu(x)\right| \le \sup_{(x,\theta)\in\mathcal{X}\times\Theta}\big\|m(x,\theta)\big\| \times \lim_{n\to\infty}\big\{\|\dot{h}_{\eta_n}-\dot{h}_{\eta_0}\|_{L^2_\mu}\|h_{\eta_n}\|_{L^2_\mu}+\|h_{\eta_n}-h_{\eta_0}\|_{L^2_\mu}\|\dot{h}_{\eta_0}\|_{L^2_\mu}\big\} = 0,$$

due to the Cauchy–Schwarz inequality, $\eta\mapsto h_\eta$ being continuously Fréchet differentiable, and $\|h_\eta\|_{L^2_\mu}=1$. Next, let $\{n_k\}$ be an arbitrary subsequence, and note that since Lemma A.7 implies $(p,\eta)\mapsto\Xi(p,P_\eta)$ is upper hemicontinuous provided η is in a neighborhood of zero, there is a further subsequence $\{\theta^*_{n_{k_j}}\}$ such that $\theta^*_{n_{k_j}}\to\theta^*$ for some $\theta^*\in\Xi(p_0,P_{\eta_0})$. Along such a subsequence, we obtain, from (A.28), (A.49), and the dominated convergence theorem,

$$\text{(A.50)}\quad \lim_{j\to\infty}\nabla F\left(\int m\big(x,\theta^*_{n_{k_j}}\big)h^2_{\eta_{n_{k_j}}}(x)\,d\mu(x)\right) \times \int m\big(x,\theta^*_{n_{k_j}}\big)\dot{h}_{\eta_{n_{k_j}}}(x)h_{\eta_{n_{k_j}}}(x)\,d\mu(x) = \nabla F\left(\int m\big(x,\theta^*\big)h^2_{\eta_0}(x)\,d\mu(x)\right)\int m\big(x,\theta^*\big)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)\,d\mu(x).$$

Hence, by Lemmas A.12 and A.14 and result (A.50), the subsequence $\{n_k\}$ has a further subsequence $\{n_{k_j}\}$, with

$$\lim_{j\to\infty} \frac{\partial}{\partial\eta}\nu\big(p_{n_{k_j}}, \Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_{n_{k_j}}} = \frac{\partial}{\partial\eta}\nu\big(p_0, \Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0}. \tag{A.51}$$

Therefore, since the subsequence $\{n_k\}$ was arbitrary, result (A.51) must also hold with $\{n\}$ in place of $\{n_{k_j}\}$. We conclude that $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ is continuous, and the second claim of the lemma then follows. *Q.E.D.*

LEMMA A.16: *Let* $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$, $\mathbf{Q} \subseteq \mathbf{M}_\mu$, *and* $\mathbf{D} \equiv \{s \in L^2_\mu : s = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{Q}\}$. *If* $\mathbf{Q}$ *is open relative to* $\mathbf{M}_\mu$ *with respect to the* τ-*topology, then, for every* $Q \in \mathbf{Q}$, *the tangent space of* $\mathbf{D}$ *at* $s = \sqrt{dQ/d\mu}$ *is given by* $\dot{\mathbf{D}} = \{h \in L^2_\mu : \int h(x)s(x)\,d\mu(x) = 0\}$.

PROOF: The proof exploits a construction in Example 3.2.1 of Bickel et al. (1993). Define

$$\mathbf{T} \equiv \left\{h \in L^2_\mu : \int h(x)s(x)\,d\mu(x) = 0\right\}, \tag{A.52}$$

and note that, by Proposition 3.2.3 in Bickel et al. (1993), we have $\dot{\mathbf{D}} \subseteq \mathbf{T}$. For the reverse inclusion, pick $h \in \mathbf{T}$ and let $\Psi : \mathbf{R} \to (0, \infty)$ be continuously differentiable, with $\Psi(0) = \Psi'(0) = 1$ and Ψ, Ψ', and Ψ'/Ψ bounded. For $s \equiv \sqrt{dQ/d\mu}$, define a parametric family of distributions to be pointwise given by

$$\begin{aligned} h^2_\eta(x) &\equiv b(\eta)s^2(x)\Psi\left(\frac{2\eta h(x)}{s(x)}\right), \\ b(\eta) &\equiv \left[\int \Psi\left(\frac{2\eta h(x)}{s(x)}\right) dQ(x)\right]^{-1}. \end{aligned} \tag{A.53}$$

Employing Proposition 2.1.1 in Bickel et al. (1993), it is straightforward to verify that $\eta \mapsto h_\eta$ is a curve in L^2_μ such that $h_0 = s$. Further note that since $\mathbf{Q}$ is open relative to $\mathbf{M}_\mu$, there exists a neighborhood $N(Q) \subseteq \mathbf{M}$ in the τ-topology such that $N(Q) \cap \mathbf{M}_\mu \subseteq \mathbf{Q}$. Let Q_η satisfy $h_\eta = \sqrt{dQ_\eta/d\mu}$ and notice that $2^{-1/2}\|h_\eta - s\|_{L^2_\mu}$ equals the Hellinger distance between Q_η and Q. Since convergence with respect to the Hellinger distance implies convergence with respect to the τ-topology, it follows that there is a neighborhood $N \subseteq \mathbf{R}$ of 0 such that $Q_\eta \in N(Q) \cap \mathbf{M}_\mu \subseteq \mathbf{Q}$ for all $\eta \in N$. We conclude $\eta \mapsto h_\eta$ is a regular parametric submodel. Moreover, by direct calculation, we also have

$$\dot{h}_0(x) = \frac{1}{2}\frac{b(0)s^2(x)\Psi'(0)2h(x)}{s(x)s(x)} + \frac{1}{2}\frac{b'(0)s^2(x)\Psi(0)}{s(x)} = h(x), \tag{A.54}$$

where we have exploited that, by the dominated convergence theorem, $b'(0) = 2\int \Psi'(0)h(x)s(x)\,d\mu(x) = 0$ due to $h \in \mathbf{T}$. Hence, from (A.54) we conclude that $h \in \dot{\mathbf{D}}$ and therefore that $\mathbf{T} = \dot{\mathbf{D}}$, which establishes the lemma. *Q.E.D.*

THEOREM A.1: *Let Assumptions* 3.2, 3.3, 3.4, 3.5 *hold and* $P \in \mathbf{P}$. *Then, the tangent space of* $\mathbf{S}$ *at* $s \equiv \sqrt{dP/d\mu}$ *is given by* $\dot{\mathbf{S}} = \{h \in L^2_\mu : \int h(x)s(x)\,d\mu(x) = 0\}$.

PROOF: The claim follows from Assumption 3.3 implying 4.2(i), Lemma A.16, and Lemmas A.2, A.8, Corollary A.3, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) implying that $\mathbf{P}$ is open in $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$. *Q.E.D.*

THEOREM A.2: *If Assumptions* 3.2, 3.3, 3.4, *and* 3.5 *hold, then the mapping* $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *pointwise defined by* $\rho(P) = \nu(\cdot, \Theta_0(P))$ *is pathwise weak-differentiable at any* $P \in \mathbf{P}$. *Moreover, for* $s \equiv \sqrt{dP/d\mu}$ *and* $\lambda(p, Q)$ *as defined in* (A.29), *the derivative* $\dot{\rho} : \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *satisfies*

$$\dot{\rho}(\dot{h}_0)(p) = 2\lambda(p, P)'\nabla F\left(\int m\big(x, \theta^*(p)\big)\,dP(x)\right) \times \int m\big(x, \theta^*(p)\big)\dot{h}_0(x)s(x)\,d\mu(x),$$

where $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *is Borel measurable and satisfies* $\theta^*(p) \in \Xi(p, P)$ (*as in* (A.24)) *for all* $p \in \mathbb{S}^{d_\theta}$.

PROOF: The existence of a Borel measurable $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ satisfying $\theta^*(p) \in \Xi(p, P)$ for all $p \in \mathbb{S}^{d_\theta}$ follows from Lemma A.13. Moreover, notice that indeed $\dot{\rho}(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ for all $\dot{h}_0 \in \dot{\mathbf{S}}$ as implied by Lemmas A.14 and A.15. We next establish that $\dot{\rho} : \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ is a continuous linear operator and then verify that it is indeed the derivative of $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$. Linearity is immediate, while continuity follows by noting that, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
\text{(A.55)}\quad & \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \big\|\dot{\rho}(\dot{h}_0)\big\|_\infty \\
&\le \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \sup_{p\in\mathbb{S}^{d_\theta}} \Big\{2\big\|\lambda(p, P)\big\| \times \sup_{v\in V_0}\big\|\nabla F(v)\big\|_F \\
&\quad \times \sqrt{d_m} \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \big\|m(x, \theta)\big\| \times \|\dot{h}_0\|_{L^2_\mu} \times \|s\|_{L^2_\mu}\Big\} \\
&< \infty,
\end{aligned}$$

where we exploited $P \in \mathbf{P}$ satisfies Assumption 3.6(iii), Lemma A.11, Assumptions 3.4(i), 3.5(ii), and $\|s\|_{L^2_\mu} = 1$.

To show that $\dot{\rho}:\dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ is the weak derivative of $\rho:\mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ at P, we need to establish that

$$\lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \left\{ \frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P))}{\eta_0} - \dot{\rho}(\dot{h}_0)(p) \right\} dB(p) = 0 \tag{A.56}$$

for all curves $\eta \mapsto P_\eta$ in $\mathbf{P}$ with $h_0 = s$ and all finite Borel measures B on $\mathbb{S}^{d_\theta}$. However, by the mean value theorem,

$$\begin{aligned} &\lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P))}{\eta_0} \, dB(p) \\ &\quad = \lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu\big(p, \Theta_0(P_\eta)\big)\bigg|_{\eta = \bar{\eta}(p, \eta_0)} dB(p) \\ &\quad = \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu\big(p, \Theta_0(P_\eta)\big)\bigg|_{\eta = 0} dB(p) = \int_{\mathbb{S}^{d_\theta}} \dot{\rho}(\dot{h}_0)(p) \, dB(p), \end{aligned} \tag{A.57}$$

where the first equality holds at each p for some $\bar{\eta}(p, \eta_0)$ a convex combination of η_0 and 0. The second equality in turn follows by Lemma A.15 justifying the use of the dominated convergence theorem, while the final equality follows by Lemma A.14 and the definition of $\dot{\rho}:\dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$. Therefore, from (A.57), (A.56) is established. *Q.E.D.*

PROOF OF THEOREM 3.2: We employ the framework in Chapter 5.2 in Bickel et al. (1993). Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{S}^{d_\theta}$, which by Corollary 14.15 in Aliprantis and Border (2006) is the dual space of $\mathbf{B}$. Let $s \equiv \sqrt{dP/d\mu}$ and $\rho:\mathbf{P} \to \mathbf{B}$ be pointwise given by $\rho(P) \equiv \nu(\cdot, \Theta_0(P))$, which has pathwise weak-derivative $\dot{\rho}$ at P by Theorem A.2. For $p \mapsto \theta^*(p)$ as in Lemma A.13 and any $B \in \mathbf{B}^*$, then let

$$\begin{aligned} \dot{\rho}^T(B)(x) \equiv \int_{\mathbb{S}^{d_\theta}} & 2\lambda(p, P)' H\big(\theta^*(p)\big) \\ & \times \big\{ m\big(x, \theta^*(p)\big) - E\big[m\big(X_i, \theta^*(p)\big)\big]\big\} s(x) \, dB(p). \end{aligned} \tag{A.58}$$

We first show that $\dot{\rho}^T:\mathbf{B}^* \to \dot{\mathbf{S}}$ is the adjoint of $\dot{\rho}:\dot{\mathbf{S}} \to \mathbf{B}$. Toward this end, we establish that: (i) $\dot{\rho}^T(B)$ is well defined for any $B \in \mathbf{B}^*$, (ii) $\dot{\rho}^T(B) \in \dot{\mathbf{S}}$, and finally (iii) $\dot{\rho}^T$ is the adjoint of $\dot{\rho}$.

By Assumption 3.4(ii), Lemma A.13, and Lemmas 4.51 and 4.52 in Aliprantis and Border (2006), the function $(x, p) \mapsto m(x, \theta^*(p))$ is jointly measurable and hence so is $p \mapsto E[m(X_i, \theta^*(p))]$. Similarly, $p \mapsto H(\theta^*(p))$ is measurable by continuity of $\theta \mapsto H(\theta)$ (see (A.28)) and Lemma A.13, while $p \mapsto \lambda(p, P)$ and $x \mapsto s(x)$ are trivially measurable by Lemma A.12 and $s \in L^2_\mu$. The joint measurability of $(p, x) \mapsto (\lambda(p, P), H(\theta^*(p)), m(x, \theta^*(p)), E[m(X_i, \theta^*(p))],$

$s(x))$ in $\mathbf{R}^{d_F} \times \mathbf{R}^{d_F \times d_m} \times \mathbf{R}^{d_m} \times \mathbf{R}^{d_m} \times \mathbf{R}$ then follows from Lemma 4.49 in Aliprantis and Border (2006), and hence

$$(p, x) \mapsto 2\lambda(p, P)' H\big(\theta^*(p)\big)\big\{m\big(x, \theta^*(p)\big) - E\big[m\big(X_i, \theta^*(p)\big)\big]\big\} s(x) \tag{A.59}$$

is jointly measurable by continuity of the composition. We conclude that $\dot{\rho}^T(B)$ is a well defined measurable function for all $B \in \mathbf{B}^*$. Moreover, for $|B|$ the total variation of B, $P \in \mathbf{P}$, Lemma A.11, and $\int s^2(x)\, d\mu(x) = 1$ imply

$$\begin{aligned} \int_{\mathcal{X}} \big(\dot{\rho}^T(B)(x)\big)^2\, d\mu(x) &\le \sup_{p \in \mathbb{S}^{d_\theta}} 16 \big\| \lambda(p, P)\big\|^2 \times \sup_{v \in V_0} \big\| \nabla F(v)\big\|_F^2 \\ &\quad \times \sup_{(x,\theta) \in \mathcal{X} \times \Theta} \big\| m(x, \theta)\big\|^2 \times |B|\big(\mathbb{S}^{d_\theta}\big) \\ &< \infty, \end{aligned} \tag{A.60}$$

which verifies $\dot{\rho}^T(B) \in L^2_\mu$ for all $B \in \mathbf{B}^*$. Similarly, since $s^2 = dP/d\mu$, exchanging the order of integration yields

$$\begin{aligned} &\int_{\mathcal{X}} \dot{\rho}^T(B)(x) s(x)\, d\mu(x) \\ &\quad = 2 \int_{\mathcal{X}} \int_{\mathbb{S}^{d_\theta}} \lambda(p, P)' H\big(\theta^*(p)\big) \\ &\qquad \times \big\{m\big(x, \theta^*(p)\big) - E\big[m\big(X_i, \theta^*(p)\big)\big]\big\}\, dB(p)\, dP(x) \\ &\quad = 0. \end{aligned} \tag{A.61}$$

Therefore, by Theorem A.1 and (A.61), we conclude that $\dot{\rho}^T(B) \in \dot{\mathbf{S}}$ for all $B \in \mathbf{B}^*$. In addition, we note that since

$$\int_{\mathbb{S}^{d_\theta}} \dot{\rho}(h)(p)\, dB(p) = \int_{\mathcal{X}} h(x) \dot{\rho}^T(B)(x)\, d\mu(x) \tag{A.62}$$

by Theorem A.1 implying $\int h(x)s(x)\, d\mu(x) = 0$ for any $h \in \dot{\mathbf{S}}$, we conclude that $\dot{\rho}^T : \mathbf{B}^* \to \dot{\mathbf{S}}$ is the adjoint of $\dot{\rho} : \dot{\mathbf{S}} \to \mathbf{B}$.

Finally, note that Theorem A.1, Theorem A.2, and Theorem 5.2.1 in Bickel et al. (1993) yield

$$\begin{aligned} &\mathrm{Cov}\bigg(\int_{\mathbb{S}^{d_\theta}} \mathbb{G}(p)\, dB_1(p), \int_{\mathbb{S}^{d_\theta}} \mathbb{G}(q)\, dB_2(q)\bigg) \\ &\quad = \frac{1}{4} \int_{\mathcal{X}} \dot{\rho}^T(B_1)(x) \dot{\rho}^T(B_2)(x)\, d\mu(x) \end{aligned} \tag{A.63}$$

$$= \int_{\mathbb{S}^{d_\theta}} \int_{\mathbb{S}^{d_\theta}} \lambda(p,P)' H\big(\theta^*(p)\big) \Omega\big(\theta^*(p), \theta^*(q)\big)$$
$$\times H\big(\theta^*(q)\big)' \lambda(q,P)\, dB_1(p)\, dB_2(q)$$

for any $B_1, B_2 \in \mathbf{B}^*$, with the second equality following from $s^2 = dP/d\mu$ and reversing the order of integration. Letting B_1 and B_2 equal the degenerate probability measures at p_1 and p_2 in (A.63) then concludes the proof. *Q.E.D.*

APPENDIX B: PROOFS OF THEOREMS 4.1, 4.2 AND COROLLARY 4.1

In this appendix, we establish Theorems 4.1 and 4.2. The proofs of Theorem 4.2 and Corollary 4.1 are self contained. The proof of Theorem 4.1, however, requires multiple steps, which we outline below.
Step 1: We first establish that $\hat{P}_n$ is consistent for P under the τ-topology (Lemma B.5), and that each neighborhood in the τ-topology contains a convex open set (Lemma B.2), which will enable us to employ the mean value theorem.
Step 2: Lemma B.3 shows that the support function is appropriately differentiable at P, which will enable us to establish that

$$\sqrt{n}\big\{\nu\big(p, \Theta_0(\hat{P}_n)\big) - \nu\big(p, \Theta_0(P)\big)\big\}$$
$$= \sqrt{n}\lambda(p, \hat{P}_{n,\tau_0(p)})' \nabla F\bigg(\int m\big(x, \tilde{\theta}(p)\big)\, d\hat{P}_{n,\tau_0(p)}(x)\bigg)$$
$$\times \int m\big(x, \tilde{\theta}(p)\big)\big(d\hat{P}_n(x) - dP(x)\big)$$

by the mean value theorem, where $\hat{P}_{n,\tau} = \tau\hat{P}_n + (1-\tau)P$, $\tau_0 : \mathbb{S}^{d_\theta} \to [0,1]$, and $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{S}^{d_\theta}$.
Step 3: In Lemma B.8, we exploit equicontinuity (Lemma B.1) to further show that, uniformly in $p \in \mathbb{S}^{d_\theta}$,

$$\sqrt{n}\lambda(p, P)' \nabla F\bigg(\int m\big(x, \tilde{\theta}(p)\big)\, d\hat{P}_{n,\tau_0(p)}(x)\bigg)$$
$$\times \int m\big(x, \tilde{\theta}(p)\big)\big(d\hat{P}_n(x) - dP(x)\big)$$
$$= \sqrt{n}\lambda(p, P)' \nabla F\bigg(\int m\big(x, \theta^*(p)\big)\, dP(x)\bigg)$$
$$\times \int m\big(x, \theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big) + o_p(1),$$

where $\theta^*(p) \in \Xi(p, P)$. A key complication is that $\Xi(p, P)$ and $\Xi(p, \hat{P}_{n,\tau_0(p)})$ may not be singleton valued. This problem is addressed employing Lemmas B.4 and B.7.
Step 4: Lemma B.9 then verifies Theorem 4.1(ii) using Steps 1, 2, and 3, and continuity of $Q \mapsto \lambda(p, Q)$. Theorem 4.1(iii) is immediate from Lemma B.9 and Lemma B.10, which shows stochastic equicontinuity.

LEMMA B.1: *Let* $\{W_i, X_i\}_{i=1}^n$ *be an i.i.d. sample with* $W_i \in \mathbf{R}$ *independent of* X_i *and* $E[W_i^2] < \infty$, *and define* $\mathcal{F} \equiv \{f : \mathcal{X} \times \mathbf{R} \to \mathbf{R} : f(x, w) = wm(x, \theta), \theta \in \Theta\}$. *If Assumptions* 3.2 *and* 3.4(ii) *hold*, *then* $\mathcal{F}$ *is Donsker*.

PROOF: For any $\theta_1, \theta_2 \in \Theta$, the Cauchy–Schwarz inequality and the mean value theorem imply that

$$\text{(B.1)} \quad \begin{aligned} &\sup_{x\in\mathcal{X}} \bigl| w\bigl(m^{(i)}(x, \theta_1) - m^{(i)}(x, \theta_2)\bigr) \bigr| \\ &\quad \le \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \bigl\| \nabla_\theta m(x, \theta) \bigr\|_F \times \|\theta_1 - \theta_2\| \times |w| = G(w)\|\theta_1 - \theta_2\|, \end{aligned}$$

where the equality holds for $G(w) \equiv M|w|$ for some constant M due to Assumption 3.4(ii). It follows that the class $\mathcal{F}$ is Lipschitz in $\theta \in \Theta$ and therefore, by Theorem 2.7.11 in van der Vaart and Wellner (1996), we conclude that

$$\text{(B.2)} \quad N_{[]}\bigl(2\varepsilon\|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2}\bigr) \le N\bigl(\varepsilon, \Theta, \|\cdot\|\bigr).$$

Letting $D = \text{diam}(\Theta)$ and $u = \varepsilon/2\|G\|_{L^2}$, a change of variables and result (B.2) then allow us to conclude that

$$\text{(B.3)} \quad \begin{aligned} &\int_0^\infty \sqrt{\log N_{[]}\bigl(\varepsilon, \mathcal{F}, \|\cdot\|_{L^2}\bigr)}\, d\varepsilon \\ &\quad = 2\|G\|_{L^2} \int_0^\infty \sqrt{\log N_{[]}\bigl(2u\|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2}\bigr)}\, du \\ &\quad \le 2\|G\|_{L^2} \int_0^\infty \sqrt{N\bigl(u, \Theta, \|\cdot\|\bigr)}\, du \\ &\quad \le 2\|G\|_{L^2} \int_0^D \sqrt{d_\theta \log(D/u)}\, du < \infty, \end{aligned}$$

where the final inequality holds due to $N(u, \Theta, \|\cdot\|) \le (\text{diam}(\Theta)/u)^{d_\theta}$. Since $\|G\|_{L^2}^2 = M^2 E[W_i^2] < \infty$, the claim of the lemma then follows from result (B.3) and Theorem 2.5.6 in van der Vaart and Wellner (1996). *Q.E.D.*

LEMMA B.2: *For any neighborhood* $N(P) \subseteq \mathbf{M}$, *there is a convex neighborhood* $N'(P) \subseteq \mathbf{M}$ *with* $N'(P) \subseteq N(P)$.

PROOF: Let $\mathbf{M}_s$ denote the set of signed, finite, countably additive Borel measures on $\mathcal{X}$ endowed with the τ-topology. Note that $\mathbf{M} \subset \mathbf{M}_s$ and that $\mathbf{M}_s$ is a topological vector space. For $\mathcal{F}$ the set of bounded scalar valued measurable functions on $\mathcal{X}$ and every $(f, \nu) \in \mathcal{F} \times \mathbf{M}_s$, define $p_f : \mathbf{M}_s \to \mathbf{R}$ by $p_f(\nu) = |\int f\, d\nu|$. The set of functionals $\{p_f\}_{f \in \mathcal{F}}$ is then a family of seminorms on $\mathbf{M}_s$ that, by Lemma 5.76(2) in Aliprantis and Border (2006), generates the τ-topology. Therefore, Theorem 5.73 in Aliprantis and Border (2006) establishes that $(\mathbf{M}_s, \tau)$ is a locally convex topological vector space. Moreover, by Lemma 2.53 in Aliprantis and Border (2006), the τ-topology in $\mathbf{M}$ is the relative topology on $\mathbf{M}$ induced by $(\mathbf{M}_s, \tau)$. Hence, letting $N^o(P)$ denote the interior of $N(P)$ (relative to $\mathbf{M}$), we obtain that $N^o(P) = N_s(P) \cap \mathbf{M}$ for some open set $N_s(P) \subseteq \mathbf{M}_s$. However, since $(\mathbf{M}_s, \tau)$ is locally convex, there exists an open (in $\mathbf{M}_s$) convex neighborhood of P with $N_s'(P) \subseteq N_s(P)$. Defining $N'(P) = N_s'(P) \cap \mathbf{M}$, we obtain the desired result by convexity of $\mathbf{M}$. *Q.E.D.*

LEMMA B.3: *Let Assumptions* 3.2, 3.3, 3.4, 3.5 *hold and* $P \in \mathbf{P}$. *For any* $Q \in \mathbf{M}$, *define* $Q_\tau \equiv \tau Q + (1-\tau)P$ *and* $\Xi(p, Q)$ *as in* (A.24). *Then, there is* $N(P) \subseteq \mathbf{M}$ *such that, for all* $(Q, p, \tau_0) \in N(P) \times \mathbb{S}^{d_\theta} \times [0, 1]$,

$$\begin{aligned}
&\frac{\partial}{\partial \tau} \nu\big(p, \Theta_0(Q_\tau)\big)\bigg|_{\tau=\tau_0} \\
&\quad = \lambda(p, Q_{\tau_0})' \nabla F\bigg(\int m\big(x, \theta^*\big)\, dQ_{\tau_0}(x)\bigg) \\
&\qquad \times \int m\big(x, \theta^*\big)\big(dQ(x) - dP(x)\big) \quad \textit{for any } \theta^* \in \Xi(p, Q_{\tau_0}).
\end{aligned}$$

PROOF: First observe that, by Lemma B.2, we may without loss of generality assume neighborhoods are convex. Hence, if $Q \in N(P)$, then $Q_\tau \in N(P)$ for all $\tau \in [0, 1]$. Since $\tau \mapsto F(\int m(x, \theta)\, dQ_\tau(x))$ is continuously differentiable in τ in a neighborhood of P by Lemma A.2 and Assumption 3.5, Lemma A.9 and Corollary 5 in Milgrom and Segal (2002) imply that, for Q in a neighborhood of P, the function $\tau \mapsto \nu(p, \Theta_0(Q_\tau))$ is directionally differentiable, with

$$\begin{aligned}
&\frac{\partial}{\partial \tau_+} \nu\big(p, \Theta_0(Q_\tau)\big)\bigg|_{\tau=\tau_0} \\
&\quad = \max_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F\bigg(\int m\big(x, \theta^*\big)\, dQ_{\tau_0}(x)\bigg) \\
&\qquad \times \int m\big(x, \theta^*\big)\big(dQ(x) - dP(x)\big),
\end{aligned} \tag{B.4}$$

$$
\text{(B.5)}\qquad \frac{\partial}{\partial \tau_-}\nu\big(p,\Theta_0(Q_\tau)\big)\Big|_{\tau=\tau_0} = \min_{\theta^*\in\Xi(p,Q_{\tau_0})} \lambda(p,Q_{\tau_0})'\nabla F\bigg(\int m\big(x,\theta^*\big)\,dQ_{\tau_0}(x)\bigg) \times \int m\big(x,\theta^*\big)\big(dQ(x)-dP(x)\big),
$$

where $\frac{\partial}{\partial\tau_+}$ and $\frac{\partial}{\partial\tau_-}$ denote the right and left derivatives, respectively. By Lemma A.10, however, for every $1\le i\le d_F$ such that $\lambda^{(i)}(p,Q_{\tau_0})\neq 0$, we must have $\theta_1^{(j)}=\theta_2^{(j)}$ for all $j\in\mathcal{S}_i$ and $\theta_1,\theta_2\in\Xi(p,Q_{\tau_0})$. Therefore, since $A\theta$ does not depend on τ, we immediately can conclude from (3), (4), and results (B.4) and (B.5) that

$$
\begin{aligned}
\text{(B.6)}\qquad & \frac{\partial}{\partial \tau_+}\nu\big(p,\Theta_0(Q_\tau)\big)\Big|_{\tau=\tau_0} \\
&= \max_{\theta^*\in\Xi(p,Q_{\tau_0})} \sum_{i:\lambda^{(i)}(p,Q_{\tau_0})\neq 0} \lambda^{(i)}(p,Q_{\tau_0}) \times \frac{\partial}{\partial\tau}F_S^{(i)}\bigg(\int m_S\big(x,\theta^*\big)\,dQ_\tau(x)\bigg)\bigg|_{\tau=\tau_0} \\
&= \min_{\theta^*\in\Xi(p,Q_{\tau_0})} \sum_{i:\lambda^{(i)}(p,Q_{\tau_0})\neq 0} \lambda^{(i)}(p,Q_{\tau_0}) \times \frac{\partial}{\partial\tau}F_S^{(i)}\bigg(\int m_S\big(x,\theta^*\big)\,dQ_\tau(x)\bigg)\bigg|_{\tau=\tau_0} \\
&= \frac{\partial}{\partial \tau_-}\nu\big(p,\Theta_0(Q_\tau)\big)\Big|_{\tau=\tau_0}.
\end{aligned}
$$

Therefore, we conclude from (B.6) that (B.4) and (B.5) agree, and the lemma follows. *Q.E.D.*

LEMMA B.4: *Let $N(P)\subseteq\mathbf{M}$ be a neighborhood of P and $\Gamma:\mathbb{S}^{d_\theta}\times N(P)\to\mathbf{R}^k$ be an upper hemicontinuous correspondence. Then, for every $\varepsilon>0$, there exists a $\delta>0$ and neighborhood $N'(P)\subseteq N(P)$ such that*

$$
\sup_{\|p-\tilde p\|<\delta}\ \sup_{Q\in N'(P)}\ \sup_{\gamma\in\Gamma(p,Q)}\ \inf_{\tilde\gamma\in\Gamma(\tilde p,P)}\|\gamma-\tilde\gamma\|<\varepsilon.
$$

PROOF: Fix $\varepsilon>0$, and, for any $\zeta>0$ and $(p,Q)\in\mathbb{S}^{d_\theta}\times N(P)$, let $\Gamma^\zeta(p,Q)\equiv\{\gamma\in\mathbf{R}^k:\inf_{\tilde\gamma\in\Gamma(p,Q)}\|\gamma-\tilde\gamma\|<\zeta\}$, and $N_\zeta(p)\equiv\{\tilde p\in\mathbb{S}^{d_\theta}:\|p-\tilde p\|<$

$\zeta\}$. Since the correspondence $\Gamma : \mathbb{S}^{d_\theta} \times N(P) \to \mathbf{R}^k$ is upper hemicontinuous, for each $p \in \mathbb{S}^{d_\theta}$ there is a $\zeta(p) > 0$ and a neighborhood $N(P|p)$ of P in $\mathbf{M}$ such that

$$\Gamma(\tilde{p}, Q) \subseteq \Gamma^{\varepsilon/2}(p, P) \tag{B.7}$$

for all $(\tilde{p}, Q) \in N_{\zeta(p)}(p) \times N(P|p)$. Since $\{N_{\zeta(p)/2}(p)\}_{p \in \mathbb{S}^{d_\theta}}$ is an open cover of $\mathbb{S}^{d_\theta}$, by compactness, there exists a finite set $\{p_i\}_{i=1}^K$ such that $\{N_{\zeta(p_i)/2}(p_i)\}_{i=1}^K$ is a subcover for $\mathbb{S}^{d_\theta}$. Further let $N'(P) \equiv N(P) \cap \{\bigcap_{i=1}^K N(P|p_i)\}$, and set $\delta \equiv \min_{1 \le i \le K} \zeta(p_i)/2$. Then note that if $p \in N_{\zeta(p_i)/2}(p_i)$ and $\|p - \tilde{p}\| < \delta$, then $p, \tilde{p} \in N_{\zeta(p_i)}(p_i)$. Therefore, since all $p \in \mathbb{S}^{d_\theta}$ satisfy $p \in N_{\zeta(p_i)/2}(p_i)$ for some $1 \le i \le K$ and $N'(P) \subseteq N(P|p_i)$ for all $1 \le i \le K$, we obtain

$$\begin{aligned}
&\sup_{\|p-\tilde{p}\|<\delta} \sup_{Q \in N'(P)} \sup_{\gamma \in \Gamma(p,Q)} \inf_{\tilde{\gamma} \in \Gamma(\tilde{p},P)} \|\gamma - \tilde{\gamma}\| \\
&\quad \le \max_{1 \le i \le K} \sup_{p,\tilde{p} \in N_{\zeta(p_i)}(p_i)} \sup_{Q \in N(P|p_i)} \sup_{\gamma \in \Gamma(p,Q)} \inf_{\tilde{\gamma} \in \Gamma(\tilde{p},P)} \|\gamma - \tilde{\gamma}\| \\
&\quad \le \max_{1 \le i \le K} \sup_{\gamma \in \Gamma^{\varepsilon/2}(p_i,P)} \inf_{\tilde{\gamma} \in \Gamma(p_i,P)} 2\|\gamma - \tilde{\gamma}\| < \varepsilon,
\end{aligned} \tag{B.8}$$

where in the second inequality we employed (B.7) and the third inequality follows by definition of $\Gamma^{\varepsilon/2}(p, P)$. *Q.E.D.*

LEMMA B.5: *Let Assumption* 3.1 *hold and* P_* *denote inner probability*. *Then for every neighborhood* $N(P) \subseteq \mathbf{M}$,

$$\liminf_{n \to \infty} P_*\big(\hat{P}_n \in N(P)\big) = 1.$$

PROOF: The empirical measure $\hat{P}_n$ is not measurable in $\mathbf{M}$ with respect to the Borel σ-field generated by the τ-topology, which is why we employ inner probabilities; see Chapter 6.2 in Dembo and Zeitouni (1998). Let $\mathcal{F}$ denote the set of scalar bounded measurable functions on $\mathcal{X}$ and, for every $(f, \nu) \in \mathcal{F} \times \mathbf{M}$, define $p_f : \mathbf{M} \to \mathbf{R}$ by $p_f(\nu) \equiv \int f(x)\, d\nu(x)$. Since the τ-topology is the coarsest topology making $\nu \mapsto p_f(\nu)$ continuous for all $f \in \mathcal{F}$, it follows that, for arbitrary but finite K, $\{U_i\}_{i=1}^K$ open sets in $\mathbf{R}$, and $\{f_i\}_{i=1}^K \in \mathcal{F}$, the sets of the form

$$\bigcap_{i=1}^K \big\{Q \in \mathbf{M} : p_{f_i}(Q) \in U_i\big\} \tag{B.9}$$

constitute a base for the τ-topology. Thus, since P is in the interior of $N(P)$, there exist an integer K_0, a finite collection $\{f_i\}_{i=1}^{K_0}$, and an $\varepsilon > 0$ such that

$\bigcap_{i=1}^{K_0}\{Q \in \mathbf{M} : |\int f_i(x)(dP(x) - dQ(x))| \le \varepsilon\} \subseteq N(P)$. Hence,

$$\begin{aligned}
\text{(B.10)}\quad & \liminf_{n\to\infty} P_*\big(\hat{P}_n \in N(P)\big) \\
& \ge \liminf_{n\to\infty} P\left(\max_{1\le i\le K_0}\left|\int f_i(x)\big(d\hat{P}_n(x) - dP(x)\big)\right| \le \varepsilon\right) \\
& = 1,
\end{aligned}$$

where the final equality follows from the law of large numbers since each f_i is bounded. *Q.E.D.*

LEMMA B.6: *If Assumptions* 3.2, 3.4(i)–(ii), 3.5 *hold and* $P \in \mathbf{P}$, *then there exists a neighborhood* $N(P) \subseteq \mathbf{M}$ *of* P *such that, for any* $1 \le i \le d_F$ *and any* $\theta, \tilde{\theta} \in \Theta$ *satisfying* $\theta^{(j)} = \tilde{\theta}^{(j)}$ *for all* $j \in \mathcal{S}_i$, *it follows that*

$$\begin{aligned}
& \nabla F_S^{(i)}\left(\int m_S(x,\theta)\,dQ(x)\right) m_S(x_0,\theta) \\
& = \nabla F_S^{(i)}\left(\int m_S(x,\tilde{\theta})\,dQ(x)\right) m_S(x_0,\tilde{\theta}) \\
& \quad \textit{for all } (Q, x_0) \in N(P) \times \mathcal{X}.
\end{aligned}$$

PROOF: By Lemma A.2, there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that the set $\mathcal{R}(Q) \equiv \{\int m(x,\theta)\,dQ(x)\}_{\theta\in\Theta}$ is compact and satisfies $\mathcal{R}(Q) \subset V_0$ for all $Q \in N(P)$. Letting $\mathcal{R}(Q)^\delta \equiv \{v \in \mathbf{R}^{d_m} : \inf_{\tilde{v}\in\mathcal{R}(Q)} \|v - \tilde{v}\| < \delta\}$, it follows from V_0 being open by Assumption 3.5 that, for each $Q \in N(P)$, there exists a $\delta_0(Q) > 0$ such that $\mathcal{R}(Q)^{\delta_0(Q)} \subset V_0$. Moreover, by Assumption 3.4(i), there exists an $M < \infty$ such that $\|m(x,\theta)\| \le M$ for all $(x,\theta) \in \mathcal{X} \times \Theta$. Hence, we obtain that if $c \in \mathbf{R}$ satisfies $|1 - c| < \delta_0(Q)/M$, then $\{c\int m(x,\theta)\,dQ(x)\}_{\theta\in\Theta} \subseteq \mathcal{R}(Q)^{\delta_0(Q)} \subset V_0$. Therefore, Assumption 3.5(i) implies that, for any $Q \in N(P)$, $1 \le i \le d_F$, and $\theta, \tilde{\theta} \in \Theta$ with $\theta^{(j)} = \tilde{\theta}^{(j)}$ for all $j \in \mathcal{S}_i$,

$$\begin{aligned}
\text{(B.11)}\quad & \nabla F_S^{(i)}\left(\int m_S(x,\theta)\,dQ(x)\right) \int m_S(x,\theta)\,dQ(x) \\
& = \frac{\partial}{\partial c}\left\{F_S^{(i)}\left(c\int m_S(x,\theta)\,dQ(x)\right)\right\}\bigg|_{c=1} \\
& = \frac{\partial}{\partial c}\left\{F_S^{(i)}\left(c\int m_S(x,\tilde{\theta})\,dQ(x)\right)\right\}\bigg|_{c=1} \\
& = \nabla F_S^{(i)}\left(\int m_S(x,\tilde{\theta})\,dQ(x)\right) \int m_S(x,\tilde{\theta})\,dQ(x).
\end{aligned}$$

Next, for any $x_0 \in \mathcal{X}$, let $D_{x_0} \in \mathbf{M}$ denote the probability measure satisfying $D_{x_0}(X_i = x_0) = 1$ and define $M_\tau(Q, D_{x_0}) \equiv (1-\tau)Q + \tau D_{x_0}$. Since $M_\tau(Q, D_{x_0}) \to Q$ in the total variation metric as $\tau \to 0$, it follows from $Q \in N(P)$ and $N(P)$ being open, that there is a $\tau_0 > 0$ such that $Q' \equiv M_{\tau_0}(Q, D_{x_0}) \in N(P)$. Thus, Lemma B.2 implies $M_\tau(Q, Q') \in N(P)$ for all $\tau \in [0, 1]$, and hence, for any $1 \le i \le d_F$ and $\theta, \tilde{\theta} \in \Theta$ with $\theta^{(j)} = \tilde{\theta}^{(j)}$ for all $j \in \mathcal{S}_i$,

$$\begin{aligned}
\text{(B.12)}\quad & \tau_0 \nabla F_S^{(i)}\left(\int m_S(x,\theta)\,dQ(x)\right)\int m_S(x,\theta)\big(dD_{x_0}(x) - dQ(x)\big) \\
&= \frac{\partial}{\partial \tau}\left\{F_S^{(i)}\left(\int m_S(x,\theta)\,dM_\tau\big(Q,Q'\big)(x)\right)\right\}\bigg|_{\tau=0} \\
&= \frac{\partial}{\partial \tau}\left\{F_S^{(i)}\left(\int m_S(x,\tilde{\theta})\,dM_\tau\big(Q,Q'\big)(x)\right)\right\}\bigg|_{\tau=0} \\
&= \tau_0 \nabla F_S^{(i)}\left(\int m_S(x,\tilde{\theta})\,dQ(x)\right)\int m_S(x,\tilde{\theta})\big(dD_{x_0}(x) - dQ(x)\big).
\end{aligned}$$

Therefore, the claim of the lemma follows from $\tau_0 > 0$ and results (B.11) and (B.12). *Q.E.D.*

LEMMA B.7: *Let Assumptions* 3.2, 3.4, 3.5, *and* 4.2(i) *hold*, $P \in \mathbf{P}$, $\Xi(p, P)$ *be as in* (A.24), *and* $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *satisfy* $\theta^*(p) \in \Xi(p, P)$ *for all* $p \in \mathbb{S}^{d_\theta}$. *Then, for each* $p \in \mathbb{S}^{d_\theta}$, *there exists a map* $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ *such that*

$$\text{(B.13)}\quad \big\|\theta^*(p) - \Pi_p\theta\big\| \le \inf_{\tilde{\theta}\in\Xi(p,P)} \sqrt{d_\theta}\,\|\tilde{\theta} - \theta\|$$

for all $\theta \in \Theta$. *In addition*, *there is a neighborhood* $N(P) \subseteq \mathbf{M}$ *such that*, *for all* $(p, Q, x_0, \theta) \in \mathbb{S}^{d_\theta} \times N(P) \times \mathcal{X} \times \Theta$,

$$\begin{aligned}
\text{(B.14)}\quad & \lambda(p,P)'\nabla F_S\left(\int m_S(x,\theta)\,dQ(x)\right)m_S(x_0,\theta) \\
&= \lambda(p,P)'\nabla F_S\left(\int m_S(x,\Pi_p\theta)\,dQ(x)\right)m_S(x_0,\Pi_p\theta).
\end{aligned}$$

PROOF: We first construct the map $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$. To this end, for each $p \in \mathbb{S}^{d_\theta}$, we define the set

$$\text{(B.15)}\quad \mathcal{I}(p) \equiv \bigcup_{i:\lambda^{(i)}(p,P)\neq 0} \mathcal{S}_i,$$

and for any $\theta \in \Theta$, let $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfy $(\Pi_p \theta)^{(j)} = \theta^*(p)^{(j)}$ if $j \notin \mathcal{I}(p)$, and $(\Pi_p \theta)^{(j)} = \theta^{(j)}$ if $j \in \mathcal{I}(p)$. Then,

$$\text{(B.16)} \quad \begin{aligned} \|\theta^*(p) - \Pi_p\theta\| &\leq \max_{j\in\mathcal{I}(p)} \sqrt{d_\theta}\big|\theta^*(p)^{(j)} - (\Pi_p\theta)^{(j)}\big| \\ &\leq \inf_{\tilde{\theta}\in\Xi(p,P)} \sqrt{d_\theta}\|\tilde{\theta} - \theta\|, \end{aligned}$$

where the first inequality follows from $\theta^*(p)^{(j)} = (\Pi_p\theta)^{(j)}$ for all $j \notin \mathcal{I}(p)$, while the second inequality is the result of $\theta^*(p)^{(j)} = \tilde{\theta}^{(j)}$ for all $\tilde{\theta} \in \Xi(p, P)$ and $j \in \mathcal{I}(p)$ by Lemma A.10, and $\theta^{(j)} = (\Pi_p\theta)^{(j)}$ for all $j \in \mathcal{I}(p)$. Moreover, since for all $1 \leq i \leq d_F$ such that $\lambda^{(i)}(p, P) \neq 0$ we have $(\Pi_p\theta)^{(j)} = \theta^{(j)}$ for all $j \in \mathcal{S}_i$, it follows from Lemma B.6 that there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that, for all $(p, Q, x_0, \theta) \in \mathbb{S}^{d_\theta} \times N(P) \times \mathcal{X} \times \Theta$,

$$\text{(B.17)} \quad \begin{aligned} &\lambda(p, P)'\nabla F_S\bigg(\int m_S(x, \theta)\, dQ(x)\bigg) m_S(x_0, \theta) \\ &\quad = \sum_{i:\lambda^{(i)}(p,P)\neq 0} \lambda^{(i)}(p, P)\nabla F_S^{(i)}\bigg(\int m_S(x, \Pi_p\theta)\, dQ(x)\bigg) m_S(x_0, \Pi_p\theta) \\ &\quad = \lambda(p, P)'\nabla F_S\bigg(\int m_S(x, \Pi_p\theta)\, dQ(x)\bigg) m_S(x_0, \Pi_p\theta). \end{aligned}$$

Therefore, the claims of the lemma follow from results (B.16) and (B.17). *Q.E.D.*

LEMMA B.8: *Let* $\{W_i, X_i\}_{i=1}^n$ *be i.i.d. with* $W_i \in \mathbf{R}$ *independent of* X_i *and* $E[W_i^2] < \infty$. *Define* $\hat{P}_{n,\tau} \equiv \tau\hat{P}_n + (1 - \tau)P$ *for any* $\tau \in [0, 1]$ *and* $\Xi(p, Q)$ *as in* (A.24). *If Assumptions* 3.1, 3.2, 3.4, 3.5, 4.2(i) *hold*, $P \in \mathbf{P}$, *and* P^W *and* $\hat{P}_n^W$ *are the population and empirical measures of* (X_i, W_i), *then, uniformly in* $(p, \tau) \in \mathbb{S}^{d_\theta} \times [0, 1]$ *and* $\theta \in \Xi(p, \hat{P}_{n,\tau})$,

$$\text{(B.18)} \quad \begin{aligned} &\sqrt{n}\lambda(p, P)'\nabla F_S\bigg(\int m_S(x, \theta)\, d\hat{P}_{n,\tau}(x)\bigg) \\ &\qquad \times \int w m_S(x, \theta)\big(d\hat{P}_n^W(x, w) - dP^W(x, w)\big) \\ &\quad = \sqrt{n}\lambda(p, P)'\nabla F_S\bigg(\int m_S\big(x, \theta^*(p)\big)\, dP(x)\bigg) \\ &\qquad \times \int w m_S\big(x, \theta^*(p)\big)\big(d\hat{P}_n^W(x, w) - dP^W(x, w)\big) + o_p(1), \end{aligned}$$

where $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *is a Borel measurable mapping that satisfies* $\theta^*(p) \in \Xi(p, P)$ *for all* $p \in \mathbb{S}^{d_\theta}$.

PROOF: If $N(P) \subseteq \mathbf{M}$ is convex and $\hat{P}_n \in N(P)$, then $\hat{P}_{n,\tau} \in N(P)$ for all $\tau \in [0,1]$. Therefore, by Lemmas A.2, A.7, B.2, and B.5, we obtain that, with inner probability tending to 1, $\{\int m(x,\theta)\,d\hat{P}_{n,\tau}\}_{\theta\in\Theta} \subset V_0$ and $\Xi(p,\hat{P}_{n,\tau})$ is well defined for all $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$. Next, let $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ be as in Lemma B.7, and note that, by (B.13),

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \|\Pi_p\theta - \theta^*(p)\| \le \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \inf_{\tilde{\theta}\in\Xi(p,P)} \sqrt{d_\theta}\|\theta - \tilde{\theta}\| = o_p(1), \tag{B.19}$$

where the final result follows from Lemmas A.7, B.2, and B.5, and Lemma B.4 applied with $\Gamma(p,Q) = \Xi(p,Q)$. Moreover, since $\theta^*(p) \in \Theta_0(P)$ for all $p \in \mathbb{S}^{d_\theta}$, results (A.23) and (B.19) further imply that

$$\liminf_{n\to\infty} P\big(\Pi_p\theta \in \Theta \text{ for all } \theta \in \Xi(p,\hat{P}_{n,\tau}) \text{ and } (p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]\big) = 1. \tag{B.20}$$

Furthermore, by Lemmas B.2, B.5, and B.7, the map $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfies, uniformly in $(p,\tau,\theta) \in \mathbb{S}^{d_\theta} \times [0,1] \times \Theta$,

$$\begin{aligned} &\sqrt{n}\lambda(p,P)'\nabla F_S\Big(\int m_S(x,\theta)\,d\hat{P}_{n,\tau}(x)\Big) \\ &\quad\times \int w m_S(x,\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big) \\ &= \sqrt{n}\lambda(p,P)'\nabla F_S\Big(\int m_S(x,\Pi_p\theta)\,d\hat{P}_{n,\tau}(x)\Big) \\ &\quad\times \int w m_S(x,\Pi_p\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big) + o_p(1). \end{aligned} \tag{B.21}$$

Next, observe that by Lemmas A.2, B.2, and B.5, it follows that, for V_0 as in Assumption 3.5, we have

$$\liminf_{n\to\infty} P\Big(\int m(x,\theta)\,d\hat{P}_{n,\tau}(x) \in V_0 \text{ for all } (\theta,\tau) \in \Theta \times [0,1]\Big) = 1. \tag{B.22}$$

Assumption 3.2 and (A.3) imply $E[m_S(X_i,\cdot)]$ is uniformly continuous, and hence, by (B.19), (B.20), and Lemma B.1,

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \Big\| \int m_S(x,\Pi_p\theta)\,d\hat{P}_{n,\tau}(x) - \int m_S\big(x,\theta^*(p)\big)\,dP(x) \Big\| \tag{B.23}$$

$$\leq \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \left\| \int \big(m_S(x, \Pi_p\theta) - m_S\big(x, \theta^*(p)\big)\big)\, dP(x)\right\| + o_p(1)$$
$$= o_p(1).$$

Thus, ∇F being uniformly continuous on V_0 by Assumption 3.5(ii), (B.20), (B.22), (B.23), and Lemma A.11 imply

$$\text{(B.24)} \quad \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \left| \lambda(p,P)' \left(\nabla F_S\left(\int m_S(x,\Pi_p\theta)\, d\hat{P}_{n,\tau}(x)\right) - \nabla F_S\left(\int m_S\big(x,\theta^*(p)\big)\, dP(x)\right)\right)\right| = o_p(1).$$

In addition, also observe that Lemma B.1 allows us to conclude that

$$\text{(B.25)} \quad \sup_{\theta\in\Theta} \sqrt{n} \left\| \int wm(x,\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big)\right\| = O_p(1).$$

Therefore, from results (B.20), (B.24), and (B.25), we obtain that, uniformly in $(p,\tau)\in\mathbb{S}^{d_\theta}\times[0,1]$ and $\theta\in\Xi(p,\hat{P}_{n,\tau})$,

$$\text{(B.26)} \quad \sqrt{n}\lambda(p,P)'\nabla F_S\left(\int m_S(x,\Pi_p\theta)\, d\hat{P}_{n,\tau}(x)\right) \times \int wm_S(x,\Pi_p\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big)$$
$$= \sqrt{n}\lambda(p,P)'\nabla F_S\left(\int m_S\big(x,\theta^*(p)\big)\, dP(x)\right) \times \int wm_S(x,\Pi_p\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big) + o_p(1).$$

To conclude, we note that (B.19), (B.20), and Lemma B.1 imply that, for some deterministic sequence $\delta_n \downarrow 0$,

$$\text{(B.27)} \quad \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \sqrt{n} \left\| \int w\big(m_S(x,\Pi_p\theta) - m_S\big(x,\theta^*(p)\big)\big)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big)\right\|$$
$$\leq \sup_{\|\theta_1-\theta_2\|<\delta_n} \sqrt{n} \left\| \int w\big(m_S(x,\theta_1)\right.$$

$$- m_S(x,\theta_2)\big)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big)\bigg\| + o_p(1)$$

$$= o_p(1).$$

Moreover, note that since $P \in \mathbf{P}$ satisfies Assumption 3.6(iii), it follows from Lemma A.11 and Assumption 3.5(ii) that $\|\lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p))dP(x))\|$ is uniformly bounded in $p \in \mathbb{S}^{d_\theta}$. Hence, by (B.27) and Cauchy–Schwarz,

$$\begin{aligned}
\text{(B.28)}\quad & \sqrt{n}\lambda(p,P)'\nabla F_S\bigg(\int m_S\big(x,\theta^*(p)\big)\,dP(x)\bigg) \\
&\quad \times \int w m_S(x,\Pi_p\theta)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big) \\
&= \sqrt{n}\lambda(p,P)'\nabla F_S\bigg(\int m_S\big(x,\theta^*(p)\big)\,dP(x)\bigg) \\
&\quad \times \int w m_S\big(x,\theta^*(p)\big)\big(d\hat{P}_n^W(x,w) - dP^W(x,w)\big),
\end{aligned}$$

uniformly in $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$ and $\theta \in \Xi(p,\hat{P}_{n,\tau})$. The lemma then follows from (B.21), (B.26), and (B.28). *Q.E.D.*

LEMMA B.9: *Let Assumptions* 3.1, 3.2, 3.3, 3.4, 3.5 *hold*, $P \in \mathbf{P}$, *and* $\Xi(p,P)$ *be as in* (A.24). *Then*,

$$\begin{aligned}
&\sup_{p\in\mathbb{S}^{d_\theta}}\bigg|\sqrt{n}\bigg\{\big(\nu\big(p,\Theta_0(\hat{P}_n)\big) - \nu\big(p,\Theta_0(P)\big)\big) - \lambda(p,P)'H\big(\theta^*(p)\big) \\
&\quad \times \int m\big(x,\theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big)\bigg\}\bigg| = o_p(1),
\end{aligned}$$

where $\theta^*: \mathbb{S}^{d_\theta} \to \Theta$ *is a Borel measurable mapping satisfying* $\theta^*(p) \in \Xi(p,P)$ *for all* $p \in \mathbb{S}^{d_\theta}$.

PROOF: For every $\tau \in [0,1]$, define $\hat{P}_{n,\tau} \equiv \tau\hat{P}_n + (1-\tau)P$ and notice that $\hat{P}_{n,0} = P$ and $\hat{P}_{n,1} = \hat{P}_n$. Employing the mean value theorem, which is valid by Lemmas B.2, B.3, and B.5, we can then conclude that, uniformly in $p \in \mathbb{S}^{d_\theta}$,

$$\begin{aligned}
\text{(B.29)}\quad & \sqrt{n}\big\{\nu\big(p,\Theta_0(\hat{P}_n)\big) - \nu\big(p,\Theta_0(P)\big)\big\} \\
&= \sqrt{n}\lambda(p,\hat{P}_{n,\tau_0(p)})'\nabla F\bigg(\int m\big(x,\tilde{\theta}(p)\big)\,d\hat{P}_{n,\tau_0(p)}(x)\bigg) \\
&\quad \times \int m\big(x,\tilde{\theta}(p)\big)\big(d\hat{P}_n(x) - dP(x)\big) + o_p(1)
\end{aligned}$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \to (0,1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \to \Theta$ such that $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{S}^{d_\theta}$. Next, fix $\varepsilon > 0$ and note that by Lemmas A.9 and A.12, there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that the correspondence $(p, Q) \mapsto \lambda(p, Q)$ is upper hemicontinuous and singleton valued for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$. Applying Lemmas B.2 and B.4 with $\Gamma(p, Q) = \lambda(p, Q)$ then implies that there exists a convex neighborhood $N'(P) \subseteq N(P) \subseteq \mathbf{M}$ such that

$$\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{Q \in N'(P)} \|\lambda(p, Q) - \lambda(p, P)\| < \varepsilon. \tag{B.30}$$

Since $N'(P)$ is convex, $\hat{P}_n \in N'(P)$ implies $\hat{P}_{n,\tau} \in N'(P)$ for all $\tau \in [0, 1]$. Therefore, we are able to conclude that

$$\begin{aligned} &\liminf_{n\to\infty} P\Big(\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0,1]} \|\lambda(p, \hat{P}_{n,\tau}) - \lambda(p, P)\| < \varepsilon\Big) \\ &\quad \geq \liminf_{n\to\infty} P\big(\hat{P}_n \in N'(P)\big) = 1, \end{aligned} \tag{B.31}$$

where the final equality follows from Lemma B.5. Thus, result (B.22) and Assumption 3.5(ii), result (B.25) applied with the random variable $W_i = 1$ almost surely, and results (B.29) and (B.31) in turn imply, uniformly in $p \in \mathbb{S}^{d_\theta}$,

$$\begin{aligned} &\sqrt{n}\{\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P))\} \\ &\quad = \sqrt{n}\lambda(p, P)'\nabla F\Big(\int m(x, \tilde{\theta}(p))\, d\hat{P}_{n,\tau_0(p)}(x)\Big) \\ &\qquad \times \int m(x, \tilde{\theta}(p))\big(d\hat{P}_n(x) - dP(x)\big) + o_p(1) \\ &\quad = \sqrt{n}\lambda(p, P)'\nabla F\Big(\int m(x, \theta^*(p))\, dP(x)\Big) \\ &\qquad \times \int m(x, \theta^*(p))\big(d\hat{P}_n(x) - dP(x)\big) + o_p(1), \end{aligned} \tag{B.32}$$

where the second equality follows from (3) and $\int A\theta(d\hat{P}_n(x) - dP(x)) = 0$ for all $\theta \in \Theta$, and Lemma B.8 applied with the random variable $W_i = 1$ almost surely. *Q.E.D.*

LEMMA B.10: *Let Assumptions* 3.1, 3.2, 3.3, 3.4, 3.5 *hold*, $P \in \mathbf{P}$, $\Xi(p, P)$ *be as in* (A.24), *and* $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *satisfy* $\theta^*(p) \in \Xi(p, P)$ *for all* $p \in \mathbb{S}^{d_\theta}$. *Then the following class is Donsker in* $\mathcal{C}(\mathbb{S}^{d_\theta})$:

$$\begin{aligned} \mathcal{F} \equiv \{ & f : \mathcal{X} \to \mathbf{R} : \\ & f(x) = \lambda(p, P)' H(\theta^*(p)) m(x, \theta^*(p)) \text{ for some } p \in \mathbb{S}^{d_\theta}\}. \end{aligned}$$

PROOF: For notational simplicity, let $H_S(\theta) \equiv \nabla F_S(\int m_S(x,\theta)\,dP(x))$, $H_S^{(i)}(\theta) \equiv \nabla F_S^{(i)}(\int m_S(x,\theta)\,dP(x))$, and

$$G_n(p) \equiv \sqrt{n}\lambda(p,P)'H\big(\theta^*(p)\big)\int m\big(x,\theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big). \tag{B.33}$$

We first note that since $\lambda(\cdot,P)$, m, and $H(\cdot)$ are bounded by Lemma A.11, Assumption 3.4(i), Assumption 3.5(ii), and $P \in \mathbf{P}$ satisfying Assumption 3.6(iii), it follows from the central limit theorem that, for any $p \in \mathbb{S}^{d_\theta}$,

$$G_n(p) \xrightarrow{L} N\big(0,\sigma^2(p)\big), \tag{B.34}$$

where $\sigma^2(p) \equiv \mathrm{Var}(\lambda(p,P)'H(\theta^*(p))m(X_i,\theta^*(p)))$. Moreover, also observe that since $\int A\theta(d\hat{P}_n(x) - dP(x)) = 0$,

$$\begin{aligned} G_n(p) &= \sqrt{n}\lambda(p,P)'H_S\big(\theta^*(p)\big)\int m_S\big(x,\theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big) \\ &= \sqrt{n}\sum_{i:\lambda^{(i)}(p,P)\neq 0}\lambda^{(i)}(p,P)H_S^{(i)}\big(\theta^*(p)\big) \\ &\quad\times\int m_S\big(x,\theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big). \end{aligned} \tag{B.35}$$

Thus, result (B.35) and Lemmas A.10 and B.6 imply $G_n(p)$ is independent of how $\theta^*(p) \in \Xi(p,P)$ is selected, and hence so is the asymptotic variance $\sigma^2(p)$.

Note that, in (B.34), it was argued that $G_n(p)$ is bounded in $p \in \mathbb{S}^{d_\theta}$, while identical arguments to those in (A.50)–(A.51) show $p \mapsto G_n(p)$ is continuous with probability 1. Hence, $G_n \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely, and to establish the lemma we only need to show the asymptotic uniform equicontinuity of G_n. Equivalently, we aim to show

$$\sup_{\|p-\tilde{p}\|<\delta_n}\big|G_n(p) - G_n(\tilde{p})\big| = o_p(1), \tag{B.36}$$

for any sequence $\delta_n \downarrow 0$. First observe that compactness of $\mathbb{S}^{d_\theta}$ and Lemma A.12 imply $\lambda(\cdot,P):\mathbb{S}^{d_\theta} \to \mathbf{R}^{d_F}$ is uniformly continuous. Therefore, by Assumption 3.5(ii), $P \in \mathbf{P}$ satisfying Assumption 3.6(iii), and result (B.25),

$$\begin{aligned} \sup_{\|p-\tilde{p}\|<\delta_n}\sqrt{n}\bigg|&\big(\lambda(p,P) - \lambda(\tilde{p},P)\big)'H_S\big(\theta^*(\tilde{p})\big) \\ &\times\int m_S\big(x,\theta^*(\tilde{p})\big)\big(d\hat{P}_n(x) - dP(x)\big)\bigg| \end{aligned} \tag{B.37}$$

$$\leq \sup_{\|p-\tilde{p}\|<\delta_n} \big\|\lambda(p,P)-\lambda(\tilde{p},P)\big\| \times \sup_{v\in V_0}\big\|\nabla F(v)\big\|_F$$

$$\times \sup_{\theta\in\Theta}\left\|\sqrt{n}\int m(x,\theta)\big(d\hat{P}_n(x)-dP(x)\big)\right\|$$

$$= o_p(1).$$

Hence, by results (B.35) and (B.37), we obtain by Lemma B.7 that, for some mapping $\Pi_p:\Theta\to\mathbf{R}^{d_\theta}$ satisfying (B.14),

$$\text{(B.38)}\qquad \sup_{\|p-\tilde{p}\|<\delta_n}\big|G_n(p)-G_n(\tilde{p})\big|$$

$$\leq \sup_{\|p-\tilde{p}\|<\delta_n}\sqrt{n}\left|\lambda(p,P)'\int\big(H_S\big(\theta^*(p)\big)m_S\big(x,\theta^*(p)\big)-H_S\big(\theta^*(\tilde{p})\big)\right.$$

$$\left.\times\, m_S\big(x,\theta^*(\tilde{p})\big)\big)\big(d\hat{P}_n(x)-dP(x)\big)\right|+o_p(1)$$

$$= \sup_{\|p-\tilde{p}\|<\delta_n}\sqrt{n}\left|\lambda(p,P)'\int\big(H_S\big(\theta^*(p)\big)m_S\big(x,\theta^*(p)\big)-H_S\big(\Pi_p\theta^*(\tilde{p})\big)\right.$$

$$\left.\times\, m_S\big(x,\Pi_p\theta^*(\tilde{p})\big)\big)\big(d\hat{P}_n(x)-dP(x)\big)\right|+o_p(1).$$

Moreover, it also follows from $\Pi_p:\Theta\to\mathbf{R}^{d_\theta}$ satisfying condition (B.13), and Lemmas A.7 and B.4, that

$$\text{(B.39)}\qquad \sup_{\|p-\tilde{p}\|<\delta_n}\big\|\theta^*(p)-\Pi_p\theta^*(\tilde{p})\big\|$$

$$\leq \sup_{\|p-\tilde{p}\|<\delta_n}\ \sup_{\tilde{\theta}\in\Xi(\tilde{p},P)}\ \inf_{\theta\in\Xi(p,P)}\sqrt{d_\theta}\|\theta-\tilde{\theta}\|=o(1).$$

Therefore, results (A.23) and (B.39) imply that, for δ_n sufficiently small, $\Pi_p\theta^*(\tilde{p})\in\Theta$ for all $\tilde{p},p\in\mathbb{S}^{d_\theta}$ with $\|\tilde{p}-p\|<\delta_n$. Hence, from (B.38) and (B.39) we conclude that, for some sequence $\gamma_n\to 0$ depending on δ_n,

$$\text{(B.40)}\qquad \sup_{\|p-\tilde{p}\|<\delta_n}\big|G_n(p)-G_n(\tilde{p})\big|$$

$$\leq \sup_{p\in\mathbb{S}^{d_\theta}}\ \sup_{\|\theta-\tilde{\theta}\|<\gamma_n}\sqrt{n}\left|\lambda(p,P)'\right.$$

$$\left.\times\int\big(H_S(\theta)m_S(x,\theta)-H_S(\tilde{\theta})m_S(x,\tilde{\theta})\big)\big(d\hat{P}_n(x)-dP(x)\big)\right|$$

$$+\,o_p(1),$$

where θ, $\tilde{\theta}$ are restricted to lie in Θ. However, note $\int m(x,\cdot)\,dP(x):\Theta\to\mathbf{R}^{d_m}$ is uniformly continuous by (A.3) and Assumption 3.2, and therefore Assumption 3.5(ii) and $P\in\mathbf{P}$ satisfying Assumption 3.6(iii) imply $\theta\mapsto H_S(\theta)$ is uniformly continuous. Therefore, $\lambda(\cdot,P)$ being bounded by Lemma A.11 and result (B.25) imply

$$
\begin{aligned}
\text{(B.41)}\quad & \sup_{p\in\mathbb{S}^{d_\theta}}\sup_{\|\theta-\tilde{\theta}\|<\gamma_n}\sqrt{n}\bigg|\lambda(p,P)'\big(H_S(\theta)-H_S(\tilde{\theta})\big) \\
&\qquad\times\int m_S(x,\tilde{\theta})\big(d\hat{P}_n(x)-dP(x)\big)\bigg| \\
&\leq \sup_{p\in\mathbb{S}^{d_\theta}}\big\|\lambda(p,P)\big\|\times\sup_{\|\theta-\tilde{\theta}\|<\gamma_n}\big\|H_S(\theta)-H_S(\tilde{\theta})\big\|_F \\
&\qquad\times\sup_{\theta\in\Theta}\bigg\|\sqrt{n}\int m(x,\theta)\big(d\hat{P}_n(x)-dP(x)\big)\bigg\| \\
&= o_p(1).
\end{aligned}
$$

In turn, it also follows from $H_S(\theta)$ being uniformly bounded in $\theta\in\Theta$ due to it being continuous and Assumption 3.2, Lemma A.11 implying $\|\lambda(p,P)\|$ is uniformly bounded in $p\in\mathbb{S}^{d_\theta}$, and Lemma B.1, that

$$
\begin{aligned}
\text{(B.42)}\quad & \sup_{p\in\mathbb{S}^{d_\theta}}\sup_{\|\theta-\tilde{\theta}\|<\gamma_n}\sqrt{n}\bigg|\lambda(p,P)'H_S(\theta) \\
&\qquad\times\int\big(m_S(x,\theta)-m_S(x,\tilde{\theta})\big)\big(d\hat{P}_n(x)-dP(x)\big)\bigg| \\
&\leq \sup_{p\in\mathbb{S}^{d_\theta}}\big\|\lambda(p,P)\big\|\times\sup_{\theta\in\Theta}\big\|H_S(\theta)\big\|_F \\
&\qquad\times\sup_{\|\theta-\tilde{\theta}\|<\gamma_n}\bigg\|\sqrt{n}\int\big(m_S(x,\theta)-m_S(x,\tilde{\theta})\big)\big(d\hat{P}_n(x)-dP(x)\big)\bigg\| \\
&= o_p(1).
\end{aligned}
$$

Hence, we conclude from (B.40), (B.41), and (B.42) that (B.36) holds, which establishes the asymptotic uniform equicontinuity of G_n. In turn, because $\mathbb{S}^{d_\theta}$ is totally bounded under $\|\cdot\|$, the process G_n is asymptotically tight in $\mathcal{C}(\mathbb{S}^{d_\theta})$ by Theorem 1.5.7 in van der Vaart and Wellner (1996). The lemma then follows from the convergence of the marginals and Theorem 1.5.4, Addendum 1.5.8, and Theorem 1.3.10 in van der Vaart and Wellner (1996). *Q.E.D.*

PROOF OF THEOREM 4.1: By Lemma B.9, $\{\nu(\cdot,\Theta_0(\hat{P}_n))\}$ has an influence function $\psi:\mathcal{X}\to\mathcal{C}(\mathbb{S}^{d_\theta})$ given by

$$\psi(x)\equiv\lambda(\cdot,P)'H\big(\theta^*(\cdot)\big)\big\{m\big(x,\theta^*(\cdot)\big)-E\big[m\big(X_i,\theta^*(\cdot)\big)\big]\big\}, \tag{B.43}$$

where $\theta^*:\mathbb{S}^{d_\theta}\to\Theta$ with $\theta^*(p)\in\Xi(p,P)$, which establishes (ii). By Theorem 3.2, $x\mapsto\psi(x)$ is the efficient influence function, and hence regularity of $\{\nu(\cdot,\Theta_0(\hat{P}_n))\}$ follows from Lemma B.10 and Theorem 18.1 in Kosorok (2008), which establishes (i). The stated convergence in distribution is then immediate from Lemmas B.9 and B.10, while the limiting process having the efficient covariance kernel is a direct result of the characterization of $I^{-1}(p_1,p_2)$ obtained in Theorem 3.2, which establishes (iii). *Q.E.D.*

PROOF OF THEOREM 4.2: Since $L:\mathcal{C}(\mathbb{S}^{d_\theta})\to\mathbf{R}_+$ is a subconvex function and $\{T_n\}$ is a regular estimator, we obtain from Theorems A.1, A.2 and Proposition 5.2.1 in Bickel et al. (1993) that

$$\liminf_{n\to\infty}E\big[L\big(\sqrt{n}\big\{T_n-\nu\big(\cdot,\Theta_0(P)\big)\big\}\big)\big]\ge E\big[L(\mathbb{G}_0)\big]. \tag{B.44}$$

Next, we aim to show that $\{E[L(\sqrt{n}\{\nu(\cdot,\Theta_0(\hat{P}_n))-\nu(\cdot,\Theta_0(P))\})]\}$ attains the lower bound. Toward this end, define

$$G_n(p)\equiv\sqrt{n}\big\{\nu\big(p,\Theta_0(\hat{P}_n)\big)-\nu\big(p,\Theta_0(P)\big)\big\}, \tag{B.45}$$

and note $G_n\in\mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely. Since L is continuous on $\mathbb{D}_0\subseteq\mathcal{C}(\mathbb{S}^{d_\theta})$ and $P(\mathbb{G}_0\in\mathbb{D}_0)=1$, Theorem 4.1 and Theorem 1.3.6 in van der Vaart and Wellner (1996) imply $L(G_n)\xrightarrow{L}L(\mathbb{G}_0)$ (in $\mathbf{R}$). Hence, since $a\mapsto a\wedge C$ is continuous and bounded on $\mathbf{R}$ for any constant $C>0$, the Portmanteau theorem yields

$$\limsup_{C\uparrow\infty}\limsup_{n\to\infty}\big|E\big[L(G_n)\wedge C\big]-E\big[L(\mathbb{G}_0)\wedge C\big]\big|=0. \tag{B.46}$$

Moreover, $L(\mathbb{G}_0)\le M_0+M_1\|\mathbb{G}_0\|_\infty^\kappa$ by hypothesis, and therefore Proposition A.2.3 in van der Vaart and Wellner (1996) yields $E[L(\mathbb{G}_0)]\le M_0+M_1E[\|\mathbb{G}_0\|_\infty^\kappa]<\infty$. Therefore, by the monotone convergence theorem,

$$\limsup_{C\uparrow\infty}\big|E\big[L(\mathbb{G}_0)\big]-E\big[L(\mathbb{G}_0)\wedge C\big]\big|=0. \tag{B.47}$$

By Assumption 3.5(ii) and Lemmas A.2, A.11, and B.2, there exists a convex neighborhood $N(P)\subseteq\mathbf{M}$ such that: (i) $\nabla F(\int m(x,\theta)\,dQ(x))$ is uniformly bounded in $(\theta,Q)\in\Theta\times N(P)$; (ii) $\lambda(p,Q)$ is uniformly bounded on $(p,Q)\in\mathbb{S}^{d_\theta}\times N(P)$; and (iii) the conditions of Lemma B.3 are satisfied for all $Q\in N(P)$. For every $\tau\in[0,1]$, define $\hat{P}_{n,\tau}\equiv\tau\hat{P}_n+(1-\tau)P$, and note that if

$\hat{P}_n \in N(P)$, then (B.29) holds, so that uniformly in $p \in \mathbb{S}^{d_\theta}$,

$$G_n = \tilde{\Delta}_n, \tag{B.48}$$
$$\tilde{\Delta}_n(p) \equiv \lambda(p, \hat{P}_{n,\tau_0(p)})' \nabla F\left(\int m(x, \tilde{\theta}(p))\, d\hat{P}_{n,\tau_0(p)}(x)\right) \times \int \sqrt{n} m(x, \tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)),$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \to (0,1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \to \Theta$ with $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for $\Xi(p, Q)$ as in (A.24) (and set $\tilde{\Delta}_n = 0$ if $\hat{P}_n \notin N(P)$). By compactness of Θ, definition of $N(P)$, and m being bounded by Assumption 3.4(i), we must have

$$\max\{\|G_n\|_\infty, \|\tilde{\Delta}_n\|_\infty\} \le \sqrt{n} C_0, \tag{B.49}$$

for some $C_0 > 0$. Therefore, $L(f) \le M_0 + M_1 \|f\|_\infty^\kappa$ for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$, (B.48) holding if $\hat{P}_n \in N(P)$, and (B.49) yield

$$\begin{aligned} &\limsup_{n\to\infty} \left|E[L(G_n)] - E[L(\tilde{\Delta}_n)]\right| \\ &\quad \le \limsup_{n\to\infty} 2(M_0 + M_1 C_0^\kappa n^{\kappa/2}) P(\hat{P}_n \notin N(P)). \end{aligned} \tag{B.50}$$

However, as shown in (B.10), there exist a finite collection $\{f_j\}_{j=1}^{K_0}$ of bounded functions and an $\varepsilon > 0$ such that $\{Q \in \mathbf{M} : \max_{1\le j\le K_0} |\int f_j(x)(dQ(x) - dP(x))| \le \varepsilon\} \subseteq N(P)$. Therefore, (B.50) and Bernstein's inequality imply

$$\begin{aligned} &\limsup_{n\to\infty} \left|E[L(G_n)] - E[L(\tilde{\Delta}_n)]\right| \\ &\quad \le 2(M_0 + M_1 C_0^\kappa) \\ &\qquad \times \limsup_{n\to\infty} \sum_{j=1}^{K_0} n^{\kappa/2} P\left(\left|\int f_j(x)(d\hat{P}_n(x) - dP(x))\right| > \varepsilon\right) \\ &\quad = 0. \end{aligned} \tag{B.51}$$

From result (B.51) and applying Cauchy–Schwarz and Markov's inequalities, we can then conclude that

$$\begin{aligned} &\limsup_{n\to\infty} \left|E[L(G_n)] - E[L(G_n) \wedge C]\right| \\ &\quad = \limsup_{n\to\infty} \left|E[L(\tilde{\Delta}_n)] - E[L(\tilde{\Delta}_n) \wedge C]\right| \end{aligned} \tag{B.52}$$

$$\leq \limsup_{n\to\infty} E\big[L(\tilde{\Delta}_n)1\{L(\tilde{\Delta}_n) > C\}\big] \leq \limsup_{n\to\infty} \frac{1}{C} E\big[L^2(\tilde{\Delta}_n)\big].$$

By construction of $N(P)$, there exists a compact set $\mathbf{C} \subset \mathbf{R}^{d_m}$ such that $\lambda(p, Q)'\nabla F(\int m(x, \theta)\, dQ(x)) \in \mathbf{C}$ for all $(p, \theta, Q) \in \mathbb{S}^{d_\theta} \times \Theta \times N(P)$. Let $\mathcal{G} \equiv \{g : \mathcal{X} \to \mathbf{R} : g(x) = c'm(x, \theta)$ for some $(c, \theta) \in \mathbf{C} \times \Theta\}$, and note that by Assumption 3.4(i) and compactness of $\mathbf{C}$, there exists a $C_1 > 0$ such that $\sup_{x\in\mathcal{X}} |g(x)| \leq C_1$ for all $g \in \mathcal{G}$. Moreover, for any $(c_1, \theta_1) \in \mathbf{C} \times \Theta$ and $(c_2, \theta_2) \in \mathbf{C} \times \Theta$, we also obtain, by Assumption 3.4(i)–(ii), that

$$\text{(B.53)} \quad \sup_{x\in\mathcal{X}} \big|c_1'm(x, \theta_1) - c_2'm(x, \theta_2)\big| \\ \leq \Big\{ \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \big\|m(x, \theta)\big\| + \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \big\|\nabla_\theta m(x, \theta)\big\|_F \times \sup_{c\in\mathbf{C}} \|c\| \Big\} \\ \times \big\{\|c_1 - c_2\| + \|\theta_1 - \theta_2\|\big\},$$

and hence the class $\mathcal{G}$ is Lipschitz in $(\theta, c) \in \Theta \times \mathbf{C}$. Letting $\|\cdot\| + \|\cdot\|$ denote the sum of the Euclidean norms on $\mathbf{R}^{d_\theta}$ and $\mathbf{R}^{d_m}$, we then obtain, by Theorem 2.7.11 in van der Vaart and Wellner (1996), that

$$\text{(B.54)} \quad N_{[]}\big(2\varepsilon C_1, \mathcal{G}, \|\cdot\|_\infty\big) \leq N\big(\varepsilon, \Theta \times \mathbf{C}, \|\cdot\| + \|\cdot\|\big) \lesssim \varepsilon^{-(d_m+d_\theta)}.$$

Consequently, since $\tilde{\Delta}_n = 0$ whenever $\hat{P}_n \notin N(P)$, the inequality $L(f) \leq M_0 + M_1\|f\|_\infty^\kappa$ for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$ implies

$$\text{(B.55)} \quad \limsup_{n\to\infty} E\big[L^2(\tilde{\Delta}_n)\big] \\ \leq \limsup_{n\to\infty} \big\{2M_0^2 + 2M_1^2 E\big[\|\tilde{\Delta}_n\|_\infty^{2\kappa}\big]\big\} \\ \leq \limsup_{n\to\infty} \left\{2M_0^2 + 2M_1^2 E\left[\sup_{g\in\mathcal{G}} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(X_i) - E[g(X_i)]\}\right|^{2\kappa}\right]\right\} \\ \lesssim 2M_0^2 + \left(\int_0^1 \sqrt{1 + \log N_{[]}\big(\varepsilon C_1, \mathcal{G}, \|\cdot\|_\infty\big)}\, d\varepsilon\right)^{2\kappa},$$

where the third inequality follows from Theorem 2.14.1 in van der Vaart and Wellner (1996). Combining results (B.52), (B.54), and (B.55), we can finally obtain

$$\text{(B.56)} \quad \limsup_{C\uparrow\infty} \limsup_{n\to\infty} \big|E\big[L(G_n)\big] - E\big[L(G_n) \wedge C\big]\big| \\ \leq \limsup_{C\uparrow\infty} \limsup_{n\to\infty} \frac{1}{C} E\big[L^2(\Delta_n)\big] = 0.$$

The claim of the theorem then follows from results (B.46), (B.47), and (B.56). *Q.E.D.*

PROOF OF COROLLARY 4.1: For any convex, compact valued set K_n, Corollary 1.10 in Li, Ogura, and Kreinovich (2002) implies that

$$\sqrt{n} d_H\big(K_n, \Theta_0(P)\big) = \sqrt{n}\big\|\nu(\cdot, K_n) - \nu\big(\cdot, \Theta_0(P)\big)\big\|_\infty, \tag{B.57}$$

and in particular $\sqrt{n} d_H(\hat{\Theta}_n, \Theta_0(P)) = \sqrt{n}\|\nu(\cdot, \Theta_0(\hat{P}_n)) - \nu(\cdot, \Theta_0(P))\|_\infty$. Therefore, the claim of the corollary follows if we can verify the conditions of Theorem 4.2 under the loss function $\bar{L}: \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ given by $\bar{L}(f) = L(\|f\|_\infty)$. To this end, note $\bar{L}(f) = L(\|f\|_\infty) = L(\|-f\|_\infty) = \bar{L}(-f)$. Moreover, since $L: \mathbf{R}_+ \to \mathbf{R}_+$ is subconvex, it follows that $0 = L(0) \le L(a)$, and hence if $L(a) = c$, then, by convexity of $\{a: L(a) \le c\}$, we must have $L(\lambda a) \le c$ for all $\lambda \in [0, 1]$. In particular, it follows that $L: \mathbf{R}_+ \to \mathbf{R}_+$ is nondecreasing. Therefore, if $\bar{L}(f_1) \le c$ and $\bar{L}(f_2) \le c$, then

$$\begin{aligned} \bar{L}\big(\lambda f_1 + (1-\lambda) f_2\big) &= L\big(\big\|\lambda f_1 + (1-\lambda) f_2\big\|_\infty\big) \\ &\le L\big(\lambda\|f_1\|_\infty + (1-\lambda)\|f_2\|_\infty\big) \le c, \end{aligned} \tag{B.58}$$

where the first inequality follows from L being nondecreasing, and the second by subconvexity of L. It follows from (B.58) that $\bar{L}: \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ is subconvex. The other conditions on $\bar{L}$ have been directly assumed, and the claim of the corollary follows from Theorem 4.2. *Q.E.D.*

APPENDIX C: PROOF OF THEOREM 4.3

The proof of Theorem 4.3 proceeds by: (i) deriving the semiparametric efficiency bound, and (ii) establishing that $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$ attains the bound. The efficiency bound is derived in Theorem C.1, after verifying that $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ is pathwise weak-differentiable (Lemma C.4) and characterizing the tangent space (Lemma C.3). A key challenge in the latter is showing that P satisfying Assumption 4.1 does not affect the tangent space (Lemma C.2). The fact that $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$ attains the efficiency bound follows readily after characterizing its influence function (Lemma C.6).

Some of the derivations in this appendix are similar to those in Appendices A and B. For conciseness, we provide more succinct derivations but include references to previous instances where analogous arguments were employed.

LEMMA C.1: *Let* $\mathbf{S}_\mathbf{L} \equiv \{s \in L^2_\mu : s = \sqrt{dP/d\mu}$ *for some* $P \in \mathbf{P}_\mathbf{L}\}$, *and Assumptions* 3.2, 3.4, 3.5, *and* 4.2(i) *hold. If* $\eta \mapsto h_\eta$ *is a curve in* $\mathbf{S}_\mathbf{L}$, *then there is a neighborhood* $N \subseteq \mathbf{R}$ *of* 0 *such that, for all* $(p, \eta_0) \in \mathbb{C} \times N$, $(p, \eta_0) \mapsto \frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ *exists, satisfies* (A.42), *and is both bounded and continuous on* $\mathbb{C} \times N$.

PROOF: First note that $\mathbf{P_L} \subseteq \mathbf{P}$ implies $\mathbf{S_L} \subseteq \mathbf{S}$. Therefore, there is a neighborhood $N_1 \subseteq \mathbf{R}$ of 0 such that (A.45) and (A.46) hold for all $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N_1$. Since for any $(p, \eta_0) \in \mathbb{C} \times N_1$, $\Xi(p, P_{\eta_0})$ is a singleton due to $P_{\eta_0} \in \mathbf{P_L}$, it follows that (A.45) and (A.46) equal each other and hence $\frac{\partial}{\partial\eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ exists and is given by (A.42) for all $(p, \eta_0) \in \mathbb{C} \times N_1$. The existence of a neighborhood $N_2 \subseteq N_1$ such that $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ is uniformly bounded in $(p, \eta_0) \in \mathbb{C} \times N_2$ then follows from (A.48), Lemmas A.2 and A.11, and Assumptions 3.4(i) and 3.5(ii).

To establish continuity, note that Lemmas A.7 and A.12 imply there is a neighborhood $N \subseteq N_2 \subseteq \mathbf{R}$ such that $(p, \eta_0) \mapsto \lambda(p, P_{\eta_0})$ and $(p, \eta_0) \mapsto \Xi(p, P_{\eta_0})$ are continuous and upper hemicontinuous, respectively, on $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$. Next, let $(p_0, \eta_0) \in \mathbb{C} \times N$ and $\{(p_n, \eta_n)\}_{n=1}^{\infty}$ be a sequence such that $(p_n, \eta_n) \to (p_0, \eta_0)$ and $(p_n, \eta_n) \in \mathbb{C} \times N$ for all n. Since $(p_n, P_{\eta_n}) \in \mathbb{C} \times \mathbf{P_L}$ for all $0 \le n < \infty$, $\Xi(p_n, P_{\eta_n}) = \{\theta_n^*\}$ for some $\theta_n^* \in \Theta$ and, by upper hemicontinuity, $\theta_n^* \to \theta_0^*$ with $\Xi(p_0, P_{\eta_0}) = \{\theta_0^*\}$. Result (A.50) and continuity of $(p, P) \mapsto \lambda(p, P)$ then imply

$$\text{(C.1)} \qquad \lim_{n\to\infty} \frac{\partial}{\partial\eta}\nu\big(p_n, \Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_n} = \frac{\partial}{\partial\eta}\nu\big(p_0, \Theta_0(P_\eta)\big)\bigg|_{\eta=\eta_0},$$

due to $\frac{\partial}{\partial\eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_n}$ satisfying (A.42) for all integer $0 \le n < \infty$. *Q.E.D.*

LEMMA C.2: *If Assumptions* 3.2, 3.4, 3.5, 4.2 *hold and* $\mathbb{C}$ *is compact*, *then the following set is open in* $\mathbf{M}$:

$$\text{(C.2)} \qquad \mathbf{M_L} \equiv \{P \in \mathbf{M} : \textit{Assumptions } 3.6\text{(i)–(iv) } \textit{and } 4.1 \textit{ hold}\}.$$

PROOF: The proof is by contradiction. Suppose there exists a $P \in \mathbf{M_L}$ such that $N(P) \subsetneq \mathbf{M_L}$ for all neighborhoods $N(P) \subseteq \mathbf{M}$ of P. Let $\mathfrak{N}_P$ be the neighborhood system of P with direction $V \succeq W$ whenever $V \subseteq W$, and recall that Lemmas A.2 and A.8, Corollary A.3, and $P \in \mathbf{M_L}$ satisfying Assumption 3.6(ii) imply that the set of $P \in \mathbf{M}$ satisfying Assumptions 3.6(i)–(iv) is open in $\mathbf{M}$. Therefore, if the lemma is false, then, for $\mathfrak{A} = \mathfrak{N}_P$, there is a net $\{Q_\alpha\}_{\alpha\in\mathfrak{A}}$ with $Q_\alpha \to P$ such that, for each $\alpha \in \mathfrak{A}$: (i) Q_α satisfies Assumption 3.6(i)–(iv), and (ii) there is a $p_\alpha \in \mathbb{C}$ with $\Xi(p_\alpha, Q_\alpha)$ (as in (A.24)) not a singleton. Furthermore, by arguing as in (A.13)–(A.15), there is a $\theta_\alpha \in \Xi(p_\alpha, Q_\alpha)$ with

$$\text{(C.3)} \qquad \mathcal{A}(\theta_\alpha, Q_\alpha) = \bigcap_{\theta\in\Xi(p_\alpha, Q_\alpha)} \mathcal{A}(\theta, Q_\alpha).$$

By compactness of $\mathbb{C}$, finiteness of the number of constraints, and Lemma A.7, we can then pass to a subnet $\{Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta\in\mathfrak{B}}$ such that, for some $(p^*, \theta^*) \in \mathbb{C} \times \Xi(p^*, P)$ and a fixed set $\mathcal{C} \subseteq \{1, \ldots, d_F\}$,

$$\text{(C.4)} \qquad (Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}) \to (P, p^*, \theta^*) \quad \text{and} \quad \mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C} \quad \forall \beta \in \mathfrak{B}.$$

Next, note that Assumption 4.2(ii) implies we can partition $\{1,\ldots,d_F\}$ into $\mathcal{I}_L \equiv \{i : \mathcal{S}_i = \emptyset\}$ and $\mathcal{I}_S \equiv \{i : \mathcal{S}_i = \{1,\ldots,d_\theta\}\}$. Since Assumption 3.2 and Q_{α_β} satisfying Assumption 3.6(ii) imply $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ is convex and $F^{(i)}(\int m(x,\cdot)\,dQ_{\alpha_\beta}(x)) : \Theta \to \mathbf{R}$ is strictly convex for all $i \in \mathcal{I}_S$, $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ being nonsingleton and (C.3) yield

$$\mathcal{C} \subseteq \mathcal{I}_L. \tag{C.5}$$

Hence, by the complementary slackness condition, $\lambda^{(i)}(p_{\alpha_\beta}, Q_{\alpha_\beta}) = 0$ for all $i \in \mathcal{I}_S$. Since Theorem 8.3.1 in Luenberger (1969) implies θ_{α_β} is a maximizer of (A.29), we obtain from the first order conditions and $\mathcal{S}_i = \emptyset$, for all $i \in \mathcal{I}_L$,

$$F_A\left(\int m_A(x)\,dQ_{\alpha_\beta}(x)\right)'\lambda(p_{\alpha_\beta}, Q_{\alpha_\beta}) = -p_{\alpha_\beta}, \tag{C.6}$$

where we exploited $\theta_{\alpha_\beta} \in \Theta^o$ due to Q_{α_β} satisfying Assumption 3.6(i). Since by construction, $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C}$, we may let $\lambda^{\mathcal{C}}(p_{\alpha_\beta}, Q_{\alpha_\beta})$, $F_A^{\mathcal{C}}(\int m_A(x)\,dQ_{\alpha_\beta}(x))$, and $F_S^{\mathcal{C}}(\int m_S(x,\theta)\,dQ_{\alpha_\beta}(x))$ respectively be the $\#\mathcal{C} \times 1$ subvector of $\lambda(p_{\alpha_\beta}, Q_{\alpha_\beta})$, $\#\mathcal{C} \times d_\theta$ submatrix of $F_A(\int m_A(x)\,dQ_{\alpha_\beta}(x))$, and $\#\mathcal{C} \times 1$ subvector of $F_S(\int m_S(x,\theta)\,dQ_{\alpha_\beta}(x))$ that correspond to the constraints indexed by $\mathcal{C}$. Since $\lambda^{(i)}(p_{\alpha_\beta}, Q_{\alpha_\beta}) = 0$ for all $i \notin \mathcal{C}$ by (C.4), we then have

$$F_A^{\mathcal{C}}\left(\int m_A(x)\,dP(x)\right)'\lambda^{\mathcal{C}}\big(p^*, P\big) = -p^*, \tag{C.7}$$

by results (C.4), (C.6), and Lemmas A.5 and A.12. Moreover, note that by definition of $\mathcal{C}$, we also obtain that

$$F_A^{\mathcal{C}}\left(\int m_A(x)\,dQ_{\alpha_\beta}(x)\right)\theta_{\alpha_\beta} = -F_S^{\mathcal{C}}\left(\int m_S(x,\theta_{\alpha_\beta})\,dQ_{\alpha_\beta}(x)\right). \tag{C.8}$$

Moreover, since $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{C}$ by (C.5), (C.8) is a linear equation in θ_{α_β}, and by $Q_{\alpha_\beta} \notin \mathbf{M_L}$ satisfying Assumption 3.6(iv) we must have $\#\mathcal{C} < d_\theta$, for otherwise (C.8) would have a unique solution in θ and (C.3) would imply $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ is a singleton. Thus, while (C.4), (C.8), and Lemma A.5 imply $\mathcal{C} \subseteq \mathcal{A}(\theta^*, P)$, we may also conclude from $\#\mathcal{C} < d_\theta$ and $\Xi(p^*, P)$ being a singleton by $(p^*, P) \in \mathbb{C} \times \mathbf{M_L}$, that we also have

$$\mathcal{A}\big(\theta^*, P\big) \setminus \mathcal{C} \neq \emptyset. \tag{C.9}$$

In what follows, we aim to establish a contradiction by showing that P will not satisfy Assumption 3.6(iv) at the point $\theta^* \in \Theta_0(P)$. To this end, for notational convenience we first define the sets

$$K_i \equiv \left\{\theta \in \Theta : F^{(i)}\left(\int m(x,\theta)\,dP(x)\right) \le 0\right\}, \tag{C.10}$$

$$E_i \equiv \left\{\theta \in \Theta : F^{(i)}\left(\int m(x,\theta)\,dP(x)\right) = 0\right\}.$$

Next, note that $\Xi(p^*, P) = \{\theta^*\}$ and convexity of $F^{(i)}(\int m(x,\cdot)\,dP(x)) : \Theta \to \mathbf{R}$ for all $1 \le i \le d_F$ imply

$$\text{(C.11)}\quad \{\theta^*\} = \left\{\bigcap_{1\le i\le d_F} K_i\right\} \cap \left\{\theta \in \Theta : \langle p^*, \theta\rangle = \nu\big(p^*, \Theta_0(P)\big)\right\}$$
$$= \left\{\bigcap_{i\in\mathcal{A}(\theta^*,P)} K_i\right\} \cap \left\{\theta \in \Theta : \langle p^*, \theta\rangle = \nu\big(p^*, \Theta_0(P)\big)\right\}.$$

Moreover, also note $\mathcal{C} \subseteq \mathcal{A}(\theta^*, P)$ implies $F_A^{\mathcal{C}}(\int m_A(x)\,dP(x))\theta^* = -F_S^{\mathcal{C}}(\int m_S(x,\theta^*)\,dP(x))$, and hence, by (C.7),

$$\text{(C.12)}\quad \lambda^{\mathcal{C}}\big(p^*, P\big)' F_S^{\mathcal{C}}\left(\int m_S\big(x,\theta^*\big)\,dP(x)\right) = \langle p^*, \theta^*\rangle = \nu\big(p^*, \Theta_0(P)\big).$$

Since $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{C}$, results (C.7) and (C.12) imply $\{\bigcap_{i\in\mathcal{C}} E_i\} \subseteq \{\theta \in \Theta : \langle p^*, \theta\rangle = \nu(p^*, \Theta_0(P))\}$, which yields

$$\text{(C.13)}\quad \{\theta^*\} = \left\{\bigcap_{i\in\mathcal{A}(\theta^*,P)\setminus\mathcal{C}} K_i\right\} \cap \left\{\bigcap_{i\in\mathcal{C}} E_i\right\},$$

due to (C.9), (C.11), and $E_i \subseteq K_i$. Next, let $\iota : \mathcal{A}(\theta^*, P) \setminus \mathcal{C} \to \{1, \ldots, \#\mathcal{A}(\theta^*, P) \setminus \mathcal{C}\}$ be a bijection, and define

$$\text{(C.14)}\quad j^* \equiv \min_{1\le j\le \#\mathcal{A}(\theta^*,P)\setminus\mathcal{C}} j : \left\{\bigcap_{i\in\mathcal{A}(\theta^*,P)\setminus\mathcal{C}:\iota(i)\le j} K_i\right\} \cap \left\{\bigcap_{i\in\mathcal{C}} E_i\right\} \text{ is a singleton,}$$

where we note j^* is well defined by (C.13), and $\{\bigcap_{i\in\mathcal{C}} E_i\}$ not being singleton by $\#\mathcal{C} < d_\theta$ and $F^{(i)}(\int m(x,\cdot)\,dP(x)) : \Theta \to \mathbf{R}$ being linear for all $i \in \mathcal{C}$. Thus, from (C.10), (C.14) and setting $i^* \equiv \iota^{-1}(j^*) \in \mathcal{A}(\theta^*, P)$, we conclude[14]

$$\text{(C.15)}\quad \{\theta^*\} = \arg\min_{\theta\in\Theta}\left\{F^{(i^*)}\left(\int m(x,\theta)\,dP(x)\right) \text{ s.t.}\right.$$
$$\left.\theta \in \left\{\bigcap_{i:\iota(i)\le j^*-1} K_i\right\} \cap \left\{\bigcap_{i\in\mathcal{C}} E_i\right\}\right\}.$$

[14]Here $\{\bigcap_{i\in\emptyset} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\}$ should be understood to equal $\{\bigcap_{i\in\mathcal{C}} E_i\}$.

However, since the constraint set is not a singleton, it follows that, for each i such that $\iota(i) \leq j^* - 1$, either $F^{(i)}(\int m(x, \theta)\, dP(x))$ is linear in θ (if $i \in \mathcal{I}_L$), or $F^{(i)}(\int m(x, \theta_i)\, dP(x)) < 0$ for some $\theta_i \in \{\bigcap_{i:\iota(i)\leq j^*-1} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\}$ (if $i \in \mathcal{I}_S$). It follows that (C.15) is an ordinary convex problem satisfying a primal qualification constraint, and, by Theorem 28.2 in Rockafellar (1970), that there exist Kuhn–Tucker vectors such that

$$\{\theta^*\} = \arg\min_{\theta\in\Theta}\Big\{F^{(i^*)}\Big(\int m(x, \theta) dP(x)\Big) + \sum_{i:\iota(i)\leq j^*-1} \gamma_i F^{(i)}\Big(\int m(x, \theta)\, dP(x)\Big) + \sum_{i\in\mathcal{C}} \pi_i F^{(i)}\Big(\int m(x, \theta)\, dP(x)\Big)\Big\}. \tag{C.16}$$

Finally, we observe that since $\theta^* \in \Theta_0(P) \subseteq \Theta^o$ by Assumption 3.6(i), result (C.16) and Corollary A.1 imply

$$-\nabla_\theta F^{(i^*)}\Big(\int m(x, \theta^*)\, dP(x)\Big) = \sum_{i:\iota(i)\leq j^*-1} \gamma_i \nabla_\theta F^{(i)}\Big(\int m(x, \theta^*)\, dP(x)\Big) + \sum_{i\in\mathcal{C}} \pi_i \nabla_\theta F^{(i)}\Big(\int m(x, \theta^*)\, dP(x)\Big). \tag{C.17}$$

Thus, we reach the desired contradiction that $P \in \mathbf{M_L}$ violates Assumption 3.6(iv). *Q.E.D.*

LEMMA C.3: *If Assumptions* 3.2, 3.4, 3.5, 4.2 *hold*, $P \in \mathbf{P_L}$, $\mathbf{S_L} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_L}\}$, *and* $\mathbb{C}$ *is compact*, *then the tangent space of* $\mathbf{S_L}$ *at* $s = \sqrt{dP/d\mu}$ *is* $\dot{\mathbf{S}}_\mathbf{L} = \{h \in L^2_\mu : \int h(x)s(x)\, d\mu(x) = 0\}$.

PROOF: The claim follows immediately from Lemmas A.16 and C.2. *Q.E.D.*

LEMMA C.4: *If Assumptions* 3.2, 3.4, 3.5, 4.2(i) *hold and* $\mathbb{C}$ *is compact*, *then the mapping* $\rho_L : \mathbf{P_L} \to \mathcal{C}(\mathbb{C})$ *pointwise defined by* $\rho_L(P) = \nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ *is pathwise weak-differentiable at any* $P \in \mathbf{P_L}$. *Moreover*, *for* $s \equiv \sqrt{dP/d\mu}$, $\lambda(p, Q)$ (*as in* (A.29)), *and* $\{\theta^*(p)\} = \Xi(p, P)$ (*as in* (A.24)), *the derivative* $\dot{\rho}_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$

satisfies

$$\dot{\rho}_L(\dot{h}_0)(p) = 2\lambda(p,P)'\nabla F\Big(\int m\big(x,\theta^*(p)\big)\,dP(x)\Big) \times \int m\big(x,\theta^*(p)\big)\dot{h}_0(x)s(x)\,d\mu(x).$$

PROOF: First note $\dot{\rho}_L(\dot{h}_0) \in \mathcal{C}(\mathbb{C})$ for any $\dot{h}_0 \in \dot{\mathbf{S}}_\mathbf{L}$ by Lemma C.1. In addition, $\dot{\rho}_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ is linear, and bounded, since by Lemma A.11, $P \in \mathbf{P}_\mathbf{L}$ satisfying Assumption 3.6(iii), and Assumptions 3.4(i) and 3.5(ii), we have

$$\text{(C.18)}\quad \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \sup_{p\in\mathbb{C}} \big|\dot{\rho}_L(\dot{h}_0)(p)\big| \le \sup_{\|\dot{h}_0\|_{L^2_\mu}} \sup_{p\in\mathbb{C}} \Big\{2\big\|\lambda(p,P)\big\| \times \sup_{v\in V_0}\big\|\nabla F(v)\big\|_F \times \sqrt{d_m} \sup_{(x,\theta)\in\mathcal{X}\times\Theta}\big\|m(x,\theta)\big\| \times \|\dot{h}_0\|_{L^2_\mu} \times \|s\|_{L^2_\mu}\Big\} < \infty.$$

Finally, note that for any curve $\eta \mapsto P_\eta$ in $\mathbf{P}_\mathbf{L}$ with $h_0 = s$ and all finite Borel measures B on $\mathbb{C}$, the mean value theorem, the dominated convergence theorem, and Lemma C.1 allow us to conclude that

$$\text{(C.19)}\quad \lim_{\eta_0\to 0}\int_{\mathbb{C}}\left\{\frac{\nu(p,\Theta_0(P_{\eta_0})) - \nu(p,\Theta_0(P))}{\eta_0} - \dot{\rho}_L(\dot{h}_0)(p)\right\} dB(p) = 0$$

(see (A.57)). Since (C.19) verifies $\dot{\rho}_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ is the weak-derivative of $\rho_L : \mathbf{P}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$, the lemma follows. *Q.E.D.*

THEOREM C.1: *Let Assumptions* 3.1, 3.2, 3.4, 3.5, 4.2 *hold*, $P \in \mathbf{P}_\mathbf{L}$, *and* $\mathbb{C}$ *be compact. For each* $\theta_1, \theta_2 \in \Theta$, *let* $H(\theta_1)$ *and* $\Omega(\theta_1,\theta_2)$ *be as in Theorem* 3.2, $\{\theta^*(p)\} = \Xi(p,P)$ *(as in* (A.24)) *and define* $\rho_L : \mathbf{P}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ *by* $\rho_L(P) \equiv \nu_{|\mathbb{C}}(\cdot,\Theta_0(P))$. *The inverse information covariance functional for estimating* $\rho_L(P)$ *is then given by*

$$\text{(C.20)}\quad I^{-1}(p_1,p_2) = \lambda(p_1,P)'H\big(\theta^*(p_1)\big)\Omega\big(\theta^*(p_1),\theta^*(p_2)\big) \times H\big(\theta^*(p_2)\big)'\lambda(p_2,P).$$

PROOF: As in the proof of Theorem 3.2, we closely follow Chapter 5.2 in Bickel et al. (1993). Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{C})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{C}$, which, by Corollary 14.15 in Aliprantis and Border (2006), is the dual space of $\mathbf{B}$. For $s \equiv \sqrt{dP/d\mu}$, then define $\dot{\rho}_L^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{L}$ pointwise by

$$\dot{\rho}_L^T(B)(x) \equiv \int_{\mathbb{C}} 2\lambda(p, P)' H\big(\theta^*(p)\big) \times \big\{m\big(x, \theta^*(p)\big) - E\big[m\big(X_i, \theta^*(p)\big)\big]\big\} s(x)\, dB(p), \tag{C.21}$$

noting that the integrand is indeed measurable by arguing as in (A.59) and exploiting that $p \mapsto \theta^*(p)$ is continuous on $\mathbb{C}$ due to Lemma A.7 and $\Xi(p, P)$ being a singleton for all $p \in \mathbb{C}$ due to $P \in \mathbf{P_L}$. For any $B \in \mathbf{B}^*$, let $\Gamma(B)$ denote the finite Borel measure on $\mathbb{S}^{d_\theta}$ given by $\Gamma(B)(A) = B(A \cap \mathbb{C})$ for any Borel set $A \subseteq \mathbb{S}^{d_\theta}$. Noting that $\dot{\rho}_L^T(B) = \dot{\rho}^T(\Gamma(B))$, it then follows from Lemma C.3 and results (A.60)–(A.62) that $\dot{\rho}_L^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{L}$ is the adjoint of $\dot{\rho}_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathbf{B}$. Lemmas C.3 and C.4 and Theorem 5.2.1 in Bickel et al. (1993) then establish the theorem. *Q.E.D.*

LEMMA C.5: *Let Assumptions* 3.2, 3.4, 3.5, 4.2 *hold*, $\mathbb{C}$ *be compact*, $P \in \mathbf{P_L}$, *and* $Q_\tau \equiv \tau Q + (1-\tau)P$ *for any* $Q \in \mathbf{M}$. *Then*, *there is a* $N(P) \subseteq \mathbf{M}$ *such that*, *for all* $(Q, p, \tau_0) \in N(P) \times \mathbb{C} \times (0, 1)$,

$$\begin{aligned} &\frac{\partial}{\partial \tau} \nu\big(p, \Theta_0(Q_\tau)\big)\bigg|_{\tau=\tau_0} \\ &\quad = \lambda(p, Q_{\tau_0})' \nabla F\bigg(\int m\big(x, \theta^*\big)\, dQ_{\tau_0}(x)\bigg) \\ &\qquad \times \int m\big(x, \theta^*\big)\big(dQ(x) - dP(x)\big) \quad \textit{where} \quad \{\theta^*\} = \Xi(p, Q_{\tau_0}). \end{aligned}$$

PROOF: By Lemmas B.2 and C.2, there is a $N(P) \subseteq \mathbf{M}$ that is convex and contained in $\mathbf{M_L}$ (as in (C.2)). Hence, if $Q \in N(P) \subseteq \mathbf{M_L}$, then $Q_\tau \in \mathbf{M_L}$ for all $\tau \in (0, 1)$, which, together with Assumption 3.5, Lemma A.9, and Corollary 5 in Milgrom and Segal (2002), imply that, for any $(Q, p) \in N(P) \times \mathbb{C}$, the function $\tau \mapsto \nu(p, \Theta_0(Q_\tau))$ is directionally differentiable with right and left derivatives given by

$$\begin{aligned} &\frac{\partial}{\partial \tau_+} \nu\big(p, \Theta_0(Q_\tau)\big)\bigg|_{\tau=\tau_0} \\ &\quad = \max_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F\bigg(\int m\big(x, \theta^*\big)\, dQ_{\tau_0}(x)\bigg) \\ &\qquad \times \int m\big(x, \theta^*\big)\big(dQ(x) - dP(x)\big), \end{aligned} \tag{C.22}$$

$$\text{(C.23)} \quad \frac{\partial}{\partial \tau_-}\nu\big(p, \Theta_0(Q_\tau)\big)\bigg|_{\tau=\tau_0} = \min_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F\bigg(\int m\big(x, \theta^*\big)\, dQ_{\tau_0}(x)\bigg) \times \int m\big(x, \theta^*\big)\big(dQ(x) - dP(x)\big)$$

(see also (B.4)–(B.5)). However, since $Q_{\tau_0} \in N(P) \subseteq \mathbf{M_L}$ for all $\tau_0 \in (0, 1)$, it follows that, for any $p \in \mathbb{C}$, the correspondence $\Xi(p, Q_{\tau_0})$ is singleton valued. We conclude (C.22) and (C.23) agree, and the lemma follows. *Q.E.D.*

LEMMA C.6: *Let Assumptions* 3.1, 3.2, 3.4, 3.5, 4.2 *hold*, $\mathbb{C}$ *be compact*, $P \in \mathbf{P_L}$, *and* $\{\theta^*(p)\} = \Xi(p, P)$. *Then*,

$$\sup_{p \in \mathbb{C}} \bigg| \sqrt{n}\bigg\{\big(\nu\big(p, \Theta_0(\hat{P}_n)\big) - \nu\big(p, \Theta_0(P)\big)\big) - \lambda(p, P)' H\big(\theta^*(p)\big) \times \int m\big(x, \theta^*(p)\big)\big(d\hat{P}_n(x) - dP(x)\big)\bigg\}\bigg| = o_p(1).$$

PROOF: By Lemma B.2, we may restrict attention to convex neighborhoods, so that if $\hat{P}_n \in N(P)$, then $\hat{P}_{n,\tau} \equiv \tau \hat{P}_n + (1-\tau)P \in N(P)$ for all $\tau \in [0, 1]$. Hence, Lemmas A.7 and B.5 imply $\Xi(p, \hat{P}_{n,\tau})$ is well defined for all $\tau \in [0, 1]$ with probability tending to 1. Moreover, since $P \in \mathbf{P_L}$ implies $\Xi(p, P)$ is singleton valued for all $p \in \mathbb{C}$, we obtain

$$\text{(C.24)} \quad \liminf_{n\to\infty} P\Big(\sup_{p \in \mathbb{C}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \big\| \theta - \theta^*(p) \big\| > \varepsilon\Big) = 0$$

for any $\varepsilon > 0$, due to Lemmas A.7, B.4, and B.5. Thus, since $p \mapsto \lambda(p, P)$ and $p \mapsto H(\theta^*(p))$ are uniformly bounded on $\mathbb{C}$ by Lemma A.11, Assumption 3.5, and $P \in \mathbf{P_L}$ satisfying Assumption 3.6(iii), we obtain

$$\text{(C.25)} \quad \sup_{p \in \mathbb{C}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \bigg\| \sqrt{n}\lambda(p, P)' H\big(\theta^*(p)\big) \times \int \big(m(x, \theta) - m\big(x, \theta^*(p)\big)\big)\big(d\hat{P}_n(x) - dP(x)\big)\bigg\| = o_p(1)$$

due to result (C.24) and Lemma B.1 (see also (B.27)–(B.28)). Additionally, since Θ is compact, result (A.3) implies $\theta \mapsto \int m(x, \theta)\, dP(x)$ is uniformly continuous on Θ, and we therefore obtain from Lemma B.1 that (see also (B.23)):

$$\sup_{p\in\mathbb{C}}\sup_{\tau\in[0,1]}\sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})}\left\|\int m(x,\theta)\,d\hat{P}_{n,\tau}(x)-\int m\big(x,\theta^*(p)\big)\,dP(x)\right\| = o_p(1). \tag{C.26}$$

Further note that $\nabla F(\int m(x,\theta)\,dP(x))$ is uniformly bounded in $\theta\in\Theta$ by Assumption 3.5 and $P\in\mathbf{P_L}$ satisfying Assumption 3.6(iii), while $\lambda(p,P)$ is uniformly bounded on $\mathbb{C}$ by Lemma A.11. Therefore, $v\mapsto\nabla F(v)$ being uniformly continuous on V_0 by Assumption 3.5(ii), together with Lemmas A.2 and B.5 and results (B.31) and (C.26), yield

$$\sup_{p\in\mathbb{C}}\sup_{\tau\in[0,1]}\sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})}\bigg\|\lambda(p,\hat{P}_{n,\tau})'\nabla F\Big(\int m(x,\theta)\,d\hat{P}_{n,\tau}(x)\Big) - \lambda(p,P)'\nabla F\Big(\int m\big(x,\theta^*(p)\big)\,dP(x)\Big)\bigg\| = o_p(1). \tag{C.27}$$

Finally, employing the mean value theorem, which is valid by Lemmas B.2, B.5, and C.5, we obtain uniformly in $p\in\mathbb{C}$ that, for some $\tau_0:\mathbb{C}\to(0,1)$ and $\tilde{\theta}:\mathbb{C}\to\Theta$ with $\tilde{\theta}(p)\in\Xi(p,\hat{P}_{n,\tau_0(p)})$ for all $p\in\mathbb{C}$,

$$\begin{aligned}\sqrt{n}\big\{\nu\big(p,\Theta_0(\hat{P}_n)\big)-\nu\big(p,\Theta_0(P)\big)\big\} &= \sqrt{n}\lambda(p,\hat{P}_{n,\tau_0(p)})'\nabla F\Big(\int m\big(x,\tilde{\theta}(p)\big)\,d\hat{P}_{n,\tau_0(p)}(x)\Big)\\ &\quad\times\int m\big(x,\tilde{\theta}(p)\big)\big(d\hat{P}_n(x)-dP(x)\big)+o_p(1)\\ &= \sqrt{n}\lambda(p,P)'H\big(\theta^*(p)\big)\int m\big(x,\theta^*(p)\big)\big(d\hat{P}_n(x)-dP(x)\big)+o_p(1),\end{aligned} \tag{C.28}$$

where the second equality follows from results (B.25), (C.25), and (C.27). *Q.E.D.*

PROOF OF THEOREM 4.3: We first show the class $\mathcal{F}\equiv\{f:\mathcal{X}\to\mathbf{R}:f(x)=\lambda(p,P)'H(\theta^*(p))m(x,\theta^*(p))$ for some $p\in\mathbb{C}\}$ is Donsker in $\mathcal{C}(\mathbb{C})$. To this end, note that $p\mapsto\lambda(p,P)'H(\theta^*(p))$ and $p\mapsto\theta^*(p)$ are continuous in $p\in\mathbb{C}$ due to Lemmas A.7 and A.12, result (A.3), Assumption 3.5, and $P\in\mathbf{P_L}$ satisfying Assumption 3.6(iii). Thus, it follows from Assumption 3.4(i)–(ii) that $f\in\mathcal{F}$ are uniformly bounded, and that the empirical process belongs to $\mathcal{C}(\mathbb{C})$. Convergence of the marginals is then immediate, while, for any sequence $\delta_n\downarrow 0$, we obtain

$$\sup_{p_1,p_2\in\mathbb{C}:\|p_1-p_2\|\le\delta_n}\left|\sqrt{n}\int\big(m\big(x,\theta^*(p_1)\big) - m\big(x,\theta^*(p_2)\big)\big)\big(d\hat{P}_n(x)-dP(x)\big)\right| = o_p(1), \tag{C.29}$$

due to Lemma B.1 and continuity of $p\mapsto\theta^*(p)$ on $\mathbb{C}$. The class $\mathcal{F}$ being Donsker then follows from (C.29), Lemma B.1, and $p\mapsto\lambda(p,P)'H(\theta^*(p))$ being uniformly continuous and bounded on $\mathbb{C}$ by compactness. Theorem 18.1 in Kosorok (2008) and Lemma C.6 then imply $\{\nu_{|\mathbb{C}}(\cdot,\Theta_0(\hat{P}_n))\}$ is a regular estimator of $\nu_{|\mathbb{C}}(\cdot,\Theta_0(P))$. The theorem then follows from the influence function of $\{\nu_{|\mathbb{C}}(\cdot,\Theta_0(\hat{P}_n))\}$ being efficient by Lemma C.6 and Theorem C.1. *Q.E.D.*

APPENDIX D: PROOFS OF THEOREMS 5.1, 5.2, 5.3, AND 5.4

The proofs of all theorems in this section are self contained, and do not require auxiliary lemmas or results.

PROOF OF THEOREM 5.1: For any metric space $(\mathbb{D},\|\cdot\|_{\mathbb{D}})$, let $\mathrm{BL}_M(\mathbb{D})$ denote the set of Lipschitz real functions on $\mathbb{D}$ whose absolute value and Lipschitz constant are bounded by M. To establish the theorem, it then suffices to show

$$\sup_{f\in\mathrm{BL}_1(\mathcal{C}(\mathbb{S}^{d_\theta}))}\big|E\big[f\big(G_n^*\big)|\{X_i\}_{i=1}^n\big]-E\big[f(\mathbb{G}_0)\big]\big| = o_p(1), \tag{D.1}$$

due to Theorem 1.12.4 in van der Vaart and Wellner (1996). Toward this end, note that Lemma B.1 implies that

$$\begin{aligned}
&\sup_{p\in\mathbb{S}^{d_\theta}}\left\|\sqrt{n}\int w\left\{m\big(x,\hat{\theta}(p)\big)-\int m\big(x,\hat{\theta}(p)\big)\,d\hat{P}_n(x)\right\}d\hat{P}_n^W(x,w)\right\| \\
&\quad\le\sup_{\theta\in\Theta}\left\|\sqrt{n}\int wm(x,\theta)\,d\hat{P}_n^W(x,w)\right\| \\
&\qquad+\sup_{(x,\theta)\in(\mathcal{X}\times\Theta)}\big\|m(x,\theta)\big\|\times\left|\sqrt{n}\int w\,d\hat{P}_n^W(x,w)\right| \\
&\quad=O_p(1),
\end{aligned} \tag{D.2}$$

due to $W_i\perp X_i$, $E[W_i]=0$ by Assumption 5.1(ii), and $(x,\theta)\mapsto m(x,\theta)$ being uniformly bounded by Assumption 3.4(i). Next, let $\Pi_p:\Theta\to\mathbf{R}^{d_\theta}$ be as in Lemma B.7, and note that Lemmas B.5 and B.7 imply, uniformly in

$p \in \mathbb{S}^{d_\theta}$,

$$\begin{aligned}
\text{(D.3)} \quad & \lambda(p,P)'\nabla F_S\Big(\int m_S\big(x,\hat\theta(p)\big)\,d\hat P_n(x)\Big)\int m_S\big(x,\hat\theta(p)\big)\,d\hat P_n(x) \\
&= \lambda(p,P)'\nabla F_S\Big(\int m_S\big(x,\Pi_p\hat\theta(p)\big)\,d\hat P_n(x)\Big) \\
&\quad \times \int m_S\big(x,\Pi_p\hat\theta(p)\big)\,d\hat P_n(x) + o_p(1) \\
&= \lambda(p,P)'\nabla F_S\Big(\int m_S\big(x,\theta^*(p)\big)\,dP(x)\Big) \\
&\quad \times \int m_S\big(x,\Pi_p\hat\theta(p)\big)\,d\hat P_n(x) + o_p(1) \\
&= \lambda(p,P)'\nabla F_S\Big(\int m_S\big(x,\theta^*(p)\big)\,dP(x)\Big) \\
&\quad \times \int m_S\big(x,\theta^*(p)\big)\,dP(x) + o_p(1),
\end{aligned}$$

where the second equality follows from (B.20), Assumption 3.4(i), and (B.24), while the third equality results from Lemma A.11, Assumption 3.5(ii), $P \in \mathbf{P}$ satisfying Assumption 3.6(iii), and result (B.23). Therefore, results (B.31), Assumption 3.5(ii), Lemmas A.2 and B.5, and result (D.2) yield, uniformly in $p \in \mathbb{S}^{d_\theta}$,

$$\begin{aligned}
\text{(D.4)} \quad & \sqrt{n}\lambda(p,\hat P_n)'\nabla F\Big(\int m\big(x,\hat\theta(p)\big)\,d\hat P_n(x)\Big) \\
&\quad \times \int w\Big\{m\big(x,\hat\theta(p)\big) - \int m\big(x,\hat\theta(p)\big)\,d\hat P_n(x)\Big\}\,d\hat P_n^W(x,w) \\
&= \sqrt{n}\lambda(p,P)'\nabla F\Big(\int m\big(x,\hat\theta(p)\big)\,d\hat P_n(x)\Big) \\
&\quad \times \int w\Big\{m\big(x,\hat\theta(p)\big) - \int m\big(x,\hat\theta(p)\big)\,d\hat P_n(x)\Big\}\,d\hat P_n^W(x,w) \\
&\quad + o_p(1) \\
&= \sqrt{n}\lambda(p,P)'\nabla F\Big(\int m\big(x,\theta^*(p)\big)\,dP(x)\Big) \\
&\quad \times \int w\Big\{m\big(x,\theta^*(p)\big) - \int m\big(x,\theta^*(p)\big)\,dP(x)\Big\}\,d\hat P_n^W(x,w) \\
&\quad + o_p(1),
\end{aligned}$$

where the second equality follows from $A\theta - \int A\theta\, d\hat{P}_n(x) = 0$, $E[W_i] = 0$ and $W_i \perp X_i$ by Assumption 5.1, Lemma B.8, and result (D.3). Next, define the process $\bar{G}_n^*$ to be pointwise given by

$$\text{(D.5)} \qquad \bar{G}_n^*(p) \equiv \sqrt{n}\lambda(p, P)' H\big(\theta^*(p)\big) \times \int w\bigg\{m\big(x, \theta^*(p)\big) - \int m\big(x, \theta^*(p)\big)\, dP(x)\bigg\}\, d\hat{P}_n^W(x, w),$$

and note that arguments identical to those in (A.50)–(A.51) imply that $\bar{G}_n^* \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely. Since all $f \in \mathrm{BL}_1(\mathcal{C}(\mathbb{S}^{d_\theta}))$ are bounded and have Lipschitz constant less than or equal to 1, for any $\eta > 0$, we must have

$$\text{(D.6)} \qquad \sup_{f \in \mathrm{BL}_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} \big|E\big[f\big(\bar{G}_n^*\big) - f\big(G_n^*\big) | \{X_i\}_{i=1}^n\big]\big| \leq \eta P\big(\big\|\bar{G}_n^* - G_n^*\big\|_\infty \leq \eta | \{X_i\}_{i=1}^n\big) + 2P\big(\big\|\bar{G}_n^* - G_n^*\big\|_\infty > \eta | \{X_i\}_{i=1}^n\big).$$

However, from (D.4), it follows that $P(\|\bar{G}_n^* - G_n^*\|_\infty > \eta | \{X_i\}_{i=1}^n) = o_p(1)$, and hence, since η in (D.6) is arbitrary,

$$\text{(D.7)} \qquad \sup_{f \in \mathrm{BL}_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} \big|E\big[f\big(G_n^*\big) | \{X_i\}_{i=1}^n\big] - E\big[f\big(\bar{G}_n^*\big) | \{X_i\}_{i=1}^n\big]\big| = o_p(1).$$

To conclude, we note that by Lemma B.10 and Theorem 2.9.6 in van der Vaart and Wellner (1996), we have

$$\text{(D.8)} \qquad \sup_{f \in \mathrm{BL}_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} \big|E\big[f\big(\bar{G}_n^*\big) | \{X_i\}_{i=1}^n\big] - E\big[f(\mathbb{G}_0)\big]\big| = o_p(1),$$

and therefore results (D.7) and (D.8) verify (D.1), which establishes the claim of the theorem. *Q.E.D.*

PROOF OF THEOREM 5.2: Let $\bar{G}_n^*$ be defined as in (D.5) and note that, by (D.4), $\|\bar{G}_n^* - G_n^*\|_\infty = o_p(1)$ unconditionally. Define a mapping $\Gamma : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathcal{C}(\mathbb{S}^{d_\theta})$ pointwise by $\Gamma(f) = \Upsilon \circ f$. The continuous mapping theorem then yields

$$\text{(D.9)} \qquad \Big|\sup_{p \in \hat{\Psi}_n} \Upsilon\big(G_n^*(p)\big) - \sup_{p \in \hat{\Psi}_n} \Upsilon\big(\bar{G}_n^*(p)\big)\Big| \leq \sup_{p \in \mathbb{S}^{d_\theta}} \big|\Upsilon\big(G_n^*(p)\big) - \Upsilon\big(\bar{G}_n^*(p)\big)\big| = \big\|\Gamma\big(G_n^*\big) - \Gamma\big(\bar{G}_n^*\big)\big\|_\infty = o_p(1).$$

Next, let $\hat{p}^* \in \arg\max_{p\in\hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))$, which is well defined by Assumption 5.2(ii) and continuity of $p \mapsto \bar{G}_n^*(p)$. Letting $\Pi_{\Psi_0}\hat{p}^*$ denote the projection of $\hat{p}^*$ onto Ψ_0 and noting $\|\hat{p}^* - \Pi_{\Psi_0}\hat{p}^*\| \le d_H(\hat{\Psi}_n, \Psi_0)$, we can then obtain

$$
\begin{aligned}
\text{(D.10)}\quad & \sup_{p\in\hat{\Psi}_n} \Upsilon\big(\bar{G}_n^*(p)\big) - \sup_{p\in\Psi_0} \Upsilon\big(\bar{G}_n^*(p)\big) \\
&\le \Upsilon\big(\bar{G}_n^*(\hat{p}^*)\big) - \Upsilon\big(\bar{G}_n^*(\Pi_{\Psi_0}\hat{p}^*)\big) \\
&\le \sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} \big|\Upsilon\big(\bar{G}_n^*(p)\big) - \Upsilon\big(\bar{G}_n^*(\tilde{p})\big)\big|.
\end{aligned}
$$

Similarly, by analogous manipulations to the term $\sup_{p\in\Psi_0} \Upsilon(\bar{G}_n^*(p)) - \sup_{p\in\hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))$, we can conclude

$$
\begin{aligned}
\text{(D.11)}\quad & \Big|\sup_{p\in\hat{\Psi}_n} \Upsilon\big(\bar{G}_n^*(p)\big) - \sup_{p\in\Psi_0} \Upsilon\big(\bar{G}_n^*(p)\big)\Big| \\
&\le \sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} \big|\Upsilon\big(\bar{G}_n^*(p)\big) - \Upsilon\big(\bar{G}_n^*(\tilde{p})\big)\big|.
\end{aligned}
$$

By Assumption 5.1, Lemma B.10, and Theorem 2.9.2 in van der Vaart and Wellner (1996), $\bar{G}_n^* \stackrel{L}{\to} \bar{\mathbb{G}}$ (unconditionally) for some tight Gaussian process $\bar{\mathbb{G}}$ in $\mathcal{C}(\mathbb{S}^{d_\theta})$. Therefore, it follows that $\sup_{p\in\mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)|$ is asymptotically tight in **R**. Next, fix $\eta > 0$, $\varepsilon > 0$, and note there then is a constant $K > 0$ such that

$$
\text{(D.12)}\quad \limsup_{n\to\infty} P\Big(\sup_{p\in\mathbb{S}^{d_\theta}} \big|\bar{G}_n^*(p)\big| > K\Big) < \eta.
$$

By Assumption 5.2(i), $\Upsilon : \mathbf{R} \to \mathbf{R}$ is continuous and hence uniformly continuous on $[-K, K]$. Therefore, there is a $\delta_0 > 0$ such that $|\Upsilon(a_1) - \Upsilon(a_2)| < \varepsilon$ whenever $|a_1 - a_2| < \delta_0$ with $a_1, a_2 \in [-K, K]$. Hence, we then obtain

$$
\begin{aligned}
\text{(D.13)}\quad & \limsup_{n\to\infty} P\Big(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} \big|\Upsilon\big(\bar{G}_n^*(p)\big) - \Upsilon\big(\bar{G}_n^*(\tilde{p})\big)\big| > \varepsilon\Big) \\
&\le \limsup_{n\to\infty} P\Big(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} \big|\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})\big| > \delta_0\Big) \\
&\quad + \limsup_{n\to\infty} P\Big(\sup_{p\in\mathbb{S}^{d_\theta}} \big|\bar{G}_n^*(p)\big| > K\Big).
\end{aligned}
$$

Moreover, since the process $p \mapsto \bar{G}_n^*(p)$ is asymptotically tight in $\mathcal{C}(\mathbb{S}^{d_\theta})$ by Lemma 1.3.8 in van der Vaart and Wellner (1996), it then follows that there exists a $\gamma_0 > 0$ such that

$$\begin{aligned}
\text{(D.14)} \quad & \limsup_{n\to\infty} P\Big(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} \big|\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})\big| > \delta_0\Big) \\
& \le \limsup_{n\to\infty} P\Big(\sup_{\|p-\tilde{p}\|\le \gamma_0} \big|\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})\big| > \delta_0\Big) \\
& \quad + \limsup_{n\to\infty} P\big(d_H(\hat{\Psi}_n, \Psi_0) > \gamma_0\big) \\
& < \eta,
\end{aligned}$$

due to $d_H(\hat{\Psi}_n, \Psi_0) = o_p(1)$ by hypothesis. Since ε, η were arbitrary, combining (D.9)–(D.14), we then obtain

$$\text{(D.15)} \quad \sup_{p\in\hat{\Psi}_n} \Upsilon\big(G_n^*(p)\big) = \sup_{p\in\Psi_0} \Upsilon\big(\bar{G}_n^*(p)\big) + o_p(1).$$

Therefore, for $\mathrm{BL}_1(\mathbf{R})$ as in (D.1), arguing as in (D.7), and using Theorem 5.1 and Theorem 10.8 in Kosorok (2008):

$$\begin{aligned}
\text{(D.16)} \quad & \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \Big| E\Big[f\Big(\sup_{p\in\hat{\Psi}_n} \Upsilon\big(G_n^*(p)\big)\Big)\big|\{X_i\}_{i=1}^n\Big] - E\Big[f\Big(\sup_{p\in\Psi_0} \Upsilon\big(\mathbb{G}_0(p)\big)\Big)\Big]\Big| \\
& \le \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \Big| E\Big[f\Big(\sup_{p\in\Psi_0} \Upsilon\big(\bar{G}_n^*(p)\big)\Big)\big|\{X_i\}_{i=1}^n\Big] \\
& \quad - E\Big[f\Big(\sup_{p\in\Psi_0} \Upsilon\big(\mathbb{G}_0(p)\big)\Big)\Big]\Big| + o_p(1) \\
& = o_p(1).
\end{aligned}$$

To conclude, observe that result (D.16) together with Lemma 10.11 in Kosorok (2008) imply that

$$\text{(D.17)} \quad P\Big(\sup_{p\in\hat{\Psi}_n} \Upsilon\big(G_n^*(p)\big) \le t \big| \{X_i\}_{i=1}^n\Big) = P\Big(\sup_{p\in\Psi_0} \Upsilon\big(\mathbb{G}_0(p)\big) \le t\Big) + o_p(1)$$

for all $t \in \mathbf{R}$ that are continuity points of the cdf of $\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p))$. Moreover, since $c_{1-\alpha}$ is itself a continuity point, for any $\varepsilon > 0$ there is an $\tilde{\varepsilon} \le \varepsilon$ such that $c_{1-\alpha} \pm \tilde{\varepsilon}$ are also continuity points and, in addition,

$$P\Big(\sup_{p\in\Psi_0} Y\big(\mathbb{G}_0(p)\big) \le c_{1-\alpha} - \tilde{\varepsilon}\Big) < 1-\alpha < P\Big(\sup_{p\in\Psi_0} Y\big(\mathbb{G}_0(p)\big) \le c_{1-\alpha} + \tilde{\varepsilon}\Big), \tag{D.18}$$

due to the cdf of $\sup_{p\in\Psi_0} Y(\mathbb{G}_0(p))$ being strictly increasing at $c_{1-\alpha}$. To conclude, define the event

$$A_n \equiv \Big\{P\Big(\sup_{p\in\hat{\Psi}_n} Y\big(G_n^*(p)\big) \le c_{1-\alpha} - \tilde{\varepsilon}|\{X_i\}_{i=1}^n\Big) < 1-\alpha < P\Big(\sup_{p\in\hat{\Psi}_n} Y\big(G_n^*(p)\big) \le c_{1-\alpha} + \tilde{\varepsilon}|\{X_i\}_{i=1}^n\Big)\Big\}, \tag{D.19}$$

and observe that since $c_{1-\alpha} \pm \tilde{\varepsilon}$ are continuity points of the cdf of $\sup_{p\in\Psi_0} Y(\mathbb{G}_0(p))$, result (D.17) yields that

$$\liminf_{n\to\infty} P\big(|\hat{c}_{1-\alpha} - c_{1-\alpha}| \le \varepsilon\big) \ge \liminf_{n\to\infty} P(A_n) = 1, \tag{D.20}$$

which establishes the claim of the theorem. *Q.E.D.*

PROOF OF THEOREM 5.3: Since support functions are continuous, it follows that $\hat{\mathfrak{M}}_n(\theta) \subseteq \mathbb{S}^{d_\theta}$ is closed and bounded and therefore compact. Moreover, by Theorem 17.31 in Aliprantis and Border (2006), $\mathfrak{M}(\theta)$ is nonempty and compact valued, while Theorem 4.1 and Corollary 1.10 in Li, Ogura, and Kreinovich (2002) imply that

$$d_H\big(\Theta_0(P), \hat{\Theta}_n\big) = O_p\big(n^{-1/2}\big). \tag{D.21}$$

In turn, result (D.21) and Lemma B.10 in Kaido (2012) yield $d_H(\hat{\mathfrak{M}}_n(\theta), \mathfrak{M}(\theta)) = o_p(1)$. Therefore, Assumption 5.2 is satisfied with $\mathfrak{M}(\theta) = \Psi_0$ and $\hat{\mathfrak{M}}_n(\theta) = \hat{\Psi}_n$. Moreover, by Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998), the cdf of $\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+$ is continuous and strictly increasing except possibly at zero. However, since $\mathfrak{M}(\theta)$ is nonempty and $\mathrm{Var}\{\mathbb{G}_0(p_0)\} > 0$ for some $p_0 \in \mathfrak{M}(\theta)$ by hypothesis, we obtain that

$$P\Big(\sup_{p\in\mathfrak{M}(\theta)} \big|-\mathbb{G}_0(p)\big|_+ \le 0\Big) \le P\big(-\mathbb{G}_0(p_0) \le 0\big) = 0.5. \tag{D.22}$$

Therefore, $\alpha < 0.5$ implies that the cdf of $\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+$ is continuous and strictly increasing at $c_{1-\alpha}(\theta)$. By Theorem 5.2, it then follows that $\hat{c}_{1-\alpha}(\theta) = c_{1-\alpha}(\theta) + o_p(1)$.

Suppose $\theta \in \Theta_0(P)^o$. Then result (D.21) implies that, with probability tending to 1, $\theta \in \hat{\Theta}_n^o$. Therefore, $J_n(\theta) = 0$ with probability tending to 1, and since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta) > 0$, we conclude that

$$\text{(D.23)} \quad \liminf_{n\to\infty} P\big(J_n(\theta) \le \hat{c}_{1-\alpha}(\theta)\big) = 1.$$

Suppose, on the other hand, that $\theta \in \partial\Theta_0(P)$. Theorem 4.1 and Lemma B.9 in Kaido (2012) then imply that

$$\text{(D.24)} \quad J_n(\theta) \xrightarrow{L} \sup_{p\in\mathfrak{M}(\theta)} \big|-\mathbb{G}_0(p)\big|_+.$$

Therefore, since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta)$ and the cdf of $\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(\theta)|_+$ is continuous at $c_{1-\alpha}(\theta)$, (D.24) yields

$$\text{(D.25)} \quad \lim_{n\to\infty} P\big(J_n(\theta) \le \hat{c}_{1-\alpha}(\theta)\big) = P\Big(\sup_{p\in\mathfrak{M}(\theta)} \big|-\mathbb{G}_0(p)\big|_+ \le c_{1-\alpha}(\theta)\Big) = 1-\alpha,$$

which establishes the claim of the theorem. *Q.E.D.*

PROOF OF THEOREM 5.4: We first study the behavior of $\{\pi_n^*\}$. To this end, define the functional $\psi : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}$ to be pointwise given by $\psi(f) = \sup_{p\in\mathbb{S}^{d_\theta}}\{\nu(p,\{\theta_0\}) - f(p)\}$, and the event $A_n \equiv \{\mathrm{co}(\Theta_0(\hat{P}_n)) = \Theta_0(\hat{P}_n)\}$. By Lemmas A.6 and B.5, $P(A_n^c) = o(1)$, and hence by Theorem 11.14 in Kosorok (2008), $P_{\eta/\sqrt{n}}(A_n^c) = o(1)$. Therefore, we obtain

$$\text{(D.26)} \quad J_n(\theta_0) = \max\big\{\psi\big(\nu\big(\cdot,\Theta_0(\hat{P}_n)\big)\big), 0\big\} + o_{P_{\eta/\sqrt{n}}}(1),$$

since $J_n(\theta_0) = \max\{\psi(\nu(\cdot,\Theta_0(\hat{P}_n))), 0\}$ whenever A_n occurs. Next, note that by Lemma B.8 in Kaido (2012), the map ψ is Hadamard differentiable at $\nu(\cdot,\Theta_0(P))$ with derivative $\dot{\psi} : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}$ pointwise given by

$$\text{(D.27)} \quad \dot{\psi}(f) = -f(p_0).$$

Moreover, the Hadamard differentiability of ψ together with Theorem 4.1 and Theorem 18.6 in Kosorok (2008) imply that $\{\psi(\nu(\cdot,\Theta_0(\hat{P}_n)))\}$ is an efficient estimator for $\psi(\nu(\cdot,\Theta_0(P)))$ and hence it is regular. Let $L_{\eta/\sqrt{n}}$ denote the implied law when $X_i \sim P_{\eta/\sqrt{n}}$, and note that the functional delta method and regularity then imply

$$\text{(D.28)} \quad \sqrt{n}\big\{\psi\big(\nu\big(\cdot,\Theta_0(\hat{P}_n)\big)\big) - \psi\big(\nu\big(\cdot,\Theta_0(P_{\eta/\sqrt{n}})\big)\big)\big\} \xrightarrow{L_{\eta/\sqrt{n}}} -\mathbb{G}_0(p_0).$$

Since by Theorem 4.1 the estimator $\{\nu(\cdot,\Theta_0(\hat{P}_n))\}$ is regular and asymptotically linear, Theorem 2.1 in van der Vaart (1991) implies $\eta \mapsto \nu(\cdot,\Theta_0(P_\eta))$ is pathwise differentiable. Hence, by the chain rule, Theorem A.2, and (D.27),

$$\text{(D.29)}\quad \frac{\partial}{\partial\eta}\psi\big(\nu\big(\cdot,\Theta_0(P_\eta)\big)\big)\bigg|_{\eta=0} = -2\int \lambda(p_0,P)'H(\theta_0)m(x,\theta_0)\dot{h}_0(x)h_0(x)\,d\mu(x) = 2\int \tilde{l}(x)\dot{h}_0(x)h_0(x)\,d\mu(x),$$

where $h_\eta \equiv \sqrt{dP_\eta/d\mu}$ and the final result holds by definition of $\tilde{l}(x)$ and $\int \dot{h}_0(x)h_0(x)\,d\mu(x)=0$. Therefore,

$$\text{(D.30)}\quad \sqrt{n}\big\{\psi\big(\nu\big(\cdot,\Theta_0(\hat{P}_n)\big)\big)-\psi\big(\nu\big(\cdot,\Theta_0(P)\big)\big)\big\} \overset{L_{\eta/\sqrt{n}}}{\to} -\mathbb{G}_0(p_0)+\eta\int 2\tilde{l}(x)\dot{h}_0(x)h_0(x)\,d\mu(x),$$

due to (D.28) and (D.29). Moreover, as shown in the proof of Theorem 5.3, $\hat{c}_{1-\alpha}(\theta_0)=c_{1-\alpha}(\theta_0)+o_p(1)$ when $X_i\sim P$ and therefore, by Theorem 11.14 in Kosorok (2008), also when $X_i\sim P_{\eta/\sqrt{n}}$. Thus, exploiting result (D.26), we obtain

$$\text{(D.31)}\quad \begin{aligned}&\lim_{n\to\infty} P_{\eta/\sqrt{n}}\big(J_n(\theta_0)>\hat{c}_{1-\alpha}(\theta_0)\big)\\ &= \lim_{n\to\infty} P_{\eta/\sqrt{n}}\big(\max\big\{\psi\big(\nu\big(\cdot,\Theta_0(\hat{P}_n)\big)\big),0\big\}>c_{1-\alpha}(\theta_0)\big)\\ &= \lim_{n\to\infty} P_{\eta/\sqrt{n}}\big(\psi\big(\nu\big(\cdot,\Theta_0(\hat{P}_n)\big)\big)>c_{1-\alpha}(\theta_0)\big)\\ &= P\bigg(-\mathbb{G}_0(p_0)>c_{1-\alpha}(\theta_0)-2\eta\int \tilde{l}(x)\dot{h}_0(x)h_0(x)\,d\mu(x)\bigg),\end{aligned}$$

where the second equality follows from $c_{1-\alpha}(\theta_0)>0$ due to $\alpha<0.5$ and the last equality is a result of (D.30). Thus (D.31) verifies that $\{\pi_n^*\}$ attains the bound in (35). Moreover, if $P_\eta\in\mathbf{H}(\theta_0)$, then by (D.29), we must have

$$\text{(D.32)}\quad \int \tilde{l}(x)\dot{h}_0(x)h_0(x)\geq 0.$$

Therefore, results (D.31) and (D.32) imply that $J_n(\theta_0)$ satisfies (34) as well.

We next establish that the upper bound in (35) holds using arguments in the proof of Theorem 25.44 in van der Vaart (1999). Fix a $P_\eta\in\mathbf{H}(\theta_0)$ and $\bar{\eta}>0$ for which we aim to show the bound, and pass to a subsequence $\{n_k\}_{k=1}^\infty$ with

$$\text{(D.33)}\quad \limsup_{n\to\infty}\pi_n(P_{\bar{\eta}/\sqrt{n}})=\lim_{k\to\infty}\pi_{n_k}(P_{\bar{\eta}/\sqrt{n_k}}).$$

Further, let $\tilde{s}(x) = 2\tilde{l}(x)h_0(x)$ and $\tilde{r}(x) = \tilde{s}(x) - \dot{h}_0(x)\langle \tilde{s}, \dot{h}_0\rangle_{L^2_\mu}/\|\dot{h}_0\|^2_{L^2_\mu}$. Then, notice that, by direct calculation, we can obtain that $\tilde{s} \in \dot{\mathbf{S}}$, $\tilde{r} \in \dot{\mathbf{S}}$, and $\langle \tilde{r}, \dot{h}_0\rangle_{L^2_\mu} = 0$. Moreover, also observe that, by result (D.29), we have

$$\text{(D.34)} \qquad \langle \tilde{s}, \dot{h}_0\rangle_{L^2_\mu} = \frac{\partial}{\partial \eta}\psi\big(\nu\big(\cdot, \Theta_0(P_\eta)\big)\big)\bigg|_{\eta=0}.$$

Proceeding as in the proof of Lemma A.16, we next build an augmented model by letting $s \equiv \sqrt{dP/d\mu}$, $\Psi : \mathbf{R} \to (0, \infty)$ be continuously differentiable, with $\Psi(0) = \Psi'(0) = 1$ and Ψ, Ψ', and Ψ'/Ψ bounded, and defining

$$\text{(D.35)} \qquad q^2_{\eta,\gamma}(x) \equiv b(\eta, \gamma)s^2(x)\Psi\bigg(\frac{2}{s(x)}\big\{\eta\dot{h}_0(x) + \gamma\tilde{r}(x)\big\}\bigg),$$

$$b(\eta, \gamma) \equiv \bigg[\int \Psi\bigg(\frac{2}{s(x)}\big\{\eta\dot{h}_0(x) + \gamma\tilde{r}(x)\big\}\bigg)\,dP(x)\bigg]^{-1}.$$

For $Q_{\eta,\gamma}$ satisfying $q_{\eta,\gamma} = \sqrt{dQ_{\eta,\gamma}/d\mu}$, using Proposition 2.1.1 in Bickel et al. (1993), it is straightforward to verify that $(\eta, \gamma) \mapsto q_{\eta,\gamma}$ is then a quadratic mean differentiable model with $q_{0,0} = \sqrt{dP/d\mu}$. Moreover, Lemmas A.2, A.8, Corollary A.3, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) imply that $Q_{\eta,\gamma} \in \mathbf{P}$ for all $(\eta, \gamma) \in N$ and N a suitably small neighborhood of $(0, 0)$ in $\mathbf{R}^2$. By Theorems 12.2.3 and 13.4.1 in Lehmann and Romano (2005), it then follows that if $\|\tilde{r}\|^2_{L^2_\mu} \neq 0$, then there exists a further subsequence $\{n_{k_j}\}_{j=1}^\infty$ such that

$$\text{(D.36)} \qquad \lim_{j\to\infty} \pi_{n_{k_j}}\big(Q_{(\eta,\gamma)/\sqrt{n_{k_j}}}\big) = \pi(\eta, \gamma)$$

for all $(\eta, \gamma) \in N$, and where π is the power function of a test in a limit experiment that takes the form

$$\text{(D.37)} \qquad Z \sim N\left(\begin{bmatrix}\eta \\ \gamma\end{bmatrix}, I_0^{-1}\right), \qquad I_0 \equiv \begin{bmatrix} 4\|\dot{h}_0\|^2_{L^2_\mu} & 0 \\ 0 & 4\|\tilde{r}\|^2_{L^2_\mu}\end{bmatrix}.$$

Next, we establish that the power function π corresponds to a test that controls size for the hypothesis

$$\text{(D.38)} \qquad H_0 : \eta\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} \le 0, \qquad H_1 : \eta\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} > 0.$$

Select any $(\eta_0, \gamma_0) \in \mathbf{R}^2$ such that $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} < 0$ and define a path $t \mapsto \tilde{P}_t$ to be given by $\tilde{P}_t \equiv Q_{(-t\eta_0, -t\gamma_0)}$. Notice that $\tilde{P}_t \in \mathbf{P}$ for t small due to $Q_{(\eta,\gamma)} \in \mathbf{P}$ for all $(\eta, \gamma) \in N$. Then, as in (D.34),

$$\text{(D.39)} \qquad \frac{\partial}{\partial t}\psi\big(\nu\big(\cdot, \Theta_0(\tilde{P}_t)\big)\big)\bigg|_{t=0} = -\big\{\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu}\big\} > 0,$$

and, in addition, since at $t = 0$, $\tilde{P}_0 = P$, we have $\psi(\nu(\cdot, \Theta_0(\tilde{P}_0))) = 0$ due to $\theta_0 \in \partial\Theta_0(P)$. Thus, from (D.39) we conclude $\tilde{P}_t \in \mathbf{H}(\theta_0)$ for t in a neighborhood of zero. Noting $Q_{(\eta_0,\gamma_0)/\sqrt{n}} = \tilde{P}_{-1/\sqrt{n}}$, it follows from (34) and (D.36) that

$$\text{(D.40)}\qquad \pi(\eta_0, \gamma_0) = \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\eta_0,\gamma_0)/\sqrt{n_{k_j}}}) = \lim_{j\to\infty} \pi_{n_{k_j}}(\tilde{P}_{-1/\sqrt{n_{k_j}}}) \le \limsup_{n\to\infty} \pi_n(\tilde{P}_{-1/\sqrt{n}}) \le \alpha.$$

Since (D.40) holds for any (η_0, γ_0) such that $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} < 0$, continuity of the power function π implies it also holds for any (η_0, γ_0) with $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} = 0$. We conclude that π corresponds to a test that controls size in (D.38). Therefore, Proposition 15.2 in van der Vaart (1999) and $\tilde{s}$ being in the linear span of $\dot{h}_0$ and $\tilde{r}$ yield

$$\text{(D.41)}\qquad \pi(\eta_0, \gamma_0) \le 1 - \Phi\left(z_{1-\alpha} - \frac{\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu}}{\sigma_0}\right),$$
$$\sigma_0^2 \equiv \frac{\langle \dot{h}_0, \tilde{s}\rangle^2_{L^2_\mu}}{4\|\dot{h}_0\|^2_{L^2_\mu}} + \frac{\langle \tilde{r}, \tilde{s}\rangle^2_{L^2_\mu}}{4\|\tilde{r}\|^2_{L^2_\mu}} = \frac{\|\tilde{s}\|^2_{L^2_\mu}}{4},$$

for any (η_0, γ_0) such that $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} > 0$. Furthermore, since both $\eta \mapsto \sqrt{dP_\eta/d\mu}$ and $\eta \mapsto \sqrt{dQ_{\eta,0}/d\mu}$ are Fréchet differentiable in L^2_μ at $\eta = 0$ with derivative $\dot{h}_0$, we also have that, for any $\bar{\eta} > 0$,

$$\text{(D.42)}\qquad \limsup_{n\to\infty} \sqrt{n}\|h_{\bar{\eta}/\sqrt{n}} - q_{\bar{\eta}/\sqrt{n},0}\|_{L^2_\mu} \le \limsup_{n\to\infty} \sqrt{n}\left\{\left\|h_{\bar{\eta}/\sqrt{n}} - h_0 - \frac{\bar{\eta}}{\sqrt{n}}\dot{h}_0\right\|_{L^2_\mu} + \left\|q_{\bar{\eta}/\sqrt{n},0} - h_0 - \frac{\bar{\eta}}{\sqrt{n}}\dot{h}_0\right\|_{L^2_\mu}\right\} = 0.$$

Hence, by Theorem 13.1.4 in Lehmann and Romano (2005), $P^n_{\bar{\eta}/\sqrt{n}}$ and $Q^n_{\bar{\eta}/\sqrt{n},0}$ converge in total variation, and thus

$$\text{(D.43)}\qquad \lim_{k\to\infty} \pi_{n_k}(P_{\bar{\eta}/\sqrt{n_k}}) = \lim_{k\to\infty} \pi_{n_k}(Q_{\bar{\eta}/\sqrt{n_k},0}).$$

To conclude, observe that since $P_\eta \in \mathbf{H}(\theta_0)$, result (D.34) implies that $\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} \ge 0$. If $\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} > 0$, then $\bar{\eta} > 0$ and results (D.33), (D.36), (D.41), and (D.43) establish that

$$\begin{aligned}\text{(D.44)}\quad \limsup_{n\to\infty} \pi_n(P_{\bar{\eta}/\sqrt{n}}) &= \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\bar{\eta},0)/\sqrt{n_{k_j}}}) = \pi(\bar{\eta},0) \\ &\leq 1-\Phi\left(z_{1-\alpha} - \frac{2\bar{\eta}E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\right),\end{aligned}$$

where we have used $\sigma_0^2 = E[\mathbb{G}_0^2(p_0)]$, $\tilde{s}(x) = 2\tilde{l}(x)h_0(x)$, and $h_0^2 = dP/d\mu$. If, on the other hand, $\langle\dot{h}_0, \tilde{s}\rangle_{L^2_\mu} = 0$, then

$$\begin{aligned}\text{(D.45)}\quad \limsup_{n\to\infty} \pi_n(P_{\bar{\eta}/\sqrt{n}}) &= \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\bar{\eta},0)/\sqrt{n_{k_j}}}) = \pi(\bar{\eta},0) \\ &\leq \alpha = 1-\Phi\left(z_{1-\alpha} - \frac{2\bar{\eta}\times 0}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\right),\end{aligned}$$

due to (D.33), (D.36), (D.43) together with $\bar{\eta}\langle\dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + 0\times\langle\tilde{r}, \tilde{s}\rangle_{L^2_\mu} = 0$ and π controlling size in (D.38). Recall that we assumed $\|\tilde{r}\|_{L^2_\mu} \neq 0$ in obtaining (D.37), and hence the theorem follows from (D.44) and (D.45) whenever $\|\tilde{r}\|_{L^2_\mu} \neq 0$. The case $\|\tilde{r}\|_{L^2_\mu} = 0$ follows from the arguments in (D.36)–(D.43) applied directly to P_η (rather than $Q_{\eta,\gamma}$). *Q.E.D.*

APPENDIX E: PROOF OF THEOREM 3.3

As in the proof of Theorem 3.2, we establish Theorem 3.3 by verifying the conditions of Theorem 5.2.1 in Bickel et al. (1993), which again requires us to: (i) characterize the tangent space at P, and (ii) show that $Q \mapsto \nu(\cdot, \Theta_{0,I}(Q))$ is pathwise weak-differentiable at P. In this setting, however, both endeavors are simpler. Lemma E.1 employs Lemma A.16 to characterize the tangent space, while Lemma E.3 shows $Q \mapsto \nu(p, \Theta_{0,I}(Q))$ is pathwise weak-differentiable at P, and Lemma E.4 extends the result to show pathwise weak-differentiability of $Q \mapsto \nu(\cdot, \Theta_{0,I}(Q))$.

Subsequent to the proof of Theorem 3.3, we briefly discuss the connection between pathwise weak-differentiability in this setting, and in the moment inequalities model studied in Theorem 3.2.

LEMMA E.1: *Let Assumption* 3.7 *hold*, $P \in \mathbf{P_I}$, *and* $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_I}\}$. *Then the tangent space of* $\mathbf{S_I}$ *at* $s = \sqrt{dP/d\mu}$ *is* $\dot{\mathbf{S}}_\mathbf{I} = \{h \in L^2_\mu : \int h(x)s(x)\,d\mu(x) = 0\}$.

PROOF: Let $P \in \mathbf{P_I}$ and $\xi(\int vz' \, dP(x))$ denote the smallest singular value of the matrix $\int vz' \, dP(x)$. Since $\mathcal{X}$ is compact by Assumption 3.7(i), it follows that vz' is bounded, and hence for any net $\{Q_\alpha\}_{\alpha \in \mathfrak{A}} \subset \mathbf{M}$ with $Q_\alpha \to P$,

$$\int vz' \, dQ_\alpha(x) \to \int vz' \, dP(x). \tag{E.1}$$

Thus, since ξ is continuous under the Frobenius norm (Bhatia (1997, p. 78)) it follows from $P \in \mathbf{P_I}$ that there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that $\xi(\int vz' \, dQ(x)) > 0$ for all $Q \in N(P)$. We conclude that $\mathbf{P_I}$ is open in $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$ and the claim follows from Lemma A.16. *Q.E.D.*

LEMMA E.2: *Let Assumption* 3.7 *hold, and* $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_I}\}$. *If* $\eta \mapsto h_\eta$ *is a curve in* $\mathbf{S_I}$ *and* $h_\eta = \sqrt{dP_\eta/d\mu}$, *then there is a neighborhood* $N \subset \mathbf{R}$ *of zero, such that, for all* $\eta_0 \in N$,

$$\frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1} \bigg|_{\eta=\eta_0} = -2\Sigma(P_{\eta_0})^{-1} \left\{ \int vz' \dot{h}_{\eta_0}(x) h_{\eta_0}(x) \, d\mu(x) \right\} \Sigma(P_{\eta_0})^{-1}, \tag{E.2}$$

and in addition, $\eta_0 \mapsto \frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1}|_{\eta=\eta_0}$ *is continuous and* $\|\frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1}|_{\eta=\eta_0}\|_F$ *is uniformly bounded in* $\eta_0 \in N$.

PROOF: Recall that if $\eta \mapsto U(\eta)$ is a square matrix valued function that is invertible at $\eta = \eta_0$, then $\frac{\partial}{\partial \eta} U(\eta)^{-1}|_{\eta=\eta_0} = -U(\eta_0)^{-1} \frac{\partial}{\partial \eta} U(\eta)|_{\eta=\eta_0} U(\eta_0)^{-1}$. Hence, since $P_\eta \in \mathbf{P_I}$ implies $\Sigma(P_\eta)$ is invertible, we obtain

$$\frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1} \bigg|_{\eta=\eta_0} = -\Sigma(P_{\eta_0})^{-1} \left\{ \int 2vz' \dot{h}_{\eta_0}(x) h_{\eta_0}(x) \, d\mu(x) \right\} \Sigma(P_{\eta_0})^{-1} \tag{E.3}$$

by exploiting that vz' is bounded by Assumption 3.7(i), and arguing as in (A.43). Moreover, since $P_{\eta_0} \in \mathbf{P_I}$ by assumption, continuity of $\eta \mapsto \Sigma(P_\eta)^{-1}$ follows from (E.1) and $\|h_\eta - h_{\eta_0}\|_{L^2_\mu} = o(1)$ implying $P_\eta \to P_{\eta_0}$ in the τ-topology. Since vz' is uniformly bounded by Assumption 3.7(i), arguing as in (A.49) in turn implies that $\int 2vz' \dot{h}_{\eta_0}(x) h_{\eta_0}(x) \, d\mu(x)$ is continuous in η_0, and hence the continuity of $\eta_0 \mapsto \frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1}|_{\eta=\eta_0}$ follows from (E.3). To conclude, note that $\|\frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1}|_{\eta=0}\|_F < \infty$ due to $\|\Sigma(P_0)^{-1}\|_F < \infty$, zv' being bounded, the Cauchy–Schwarz inequality, $\|h_0\|_{L^2_\mu} = 1$, and $\|\dot{h}_0\|_{L^2_\mu} < \infty$ because $\eta \mapsto h_\eta$ is Fréchet differentiable. Hence, since $\|\frac{\partial}{\partial \eta} \Sigma(P_\eta)^{-1}|_{\eta=0}\|_F$ is finite, continuity implies it must be uniformly bounded in a neighborhood of zero, and the lemma follows. *Q.E.D.*

LEMMA E.3: *Let Assumption* 3.7 *hold, and* $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_I}\}$. *If* $\eta \mapsto h_\eta$ *is a curve in* $\mathbf{S_I}$ *and* $h_\eta = \sqrt{dP_\eta/d\mu}$, *then there is a* *neighborhood $N \subset \mathbf{R}$ of zero, such that, for all $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$,*

$$\text{(E.4)} \qquad \frac{\partial}{\partial \eta} \nu\big(p, \Theta_{0,I}(P_\eta)\big)\Big|_{\eta=\eta_0} = 2\int \big\{\psi_\nu(p, x, P_{\eta_0}) - \psi_\Sigma(p, x, P_{\eta_0})\big\}\dot{h}_{\eta_0}(x) h_{\eta_0}(x)\, d\mu(x),$$

where ψ_ν and ψ_Σ are as defined in equations (16) *and* (17), *respectively. In addition, N may be chosen so that $(p, \eta_0) \mapsto \frac{\partial}{\partial \eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ is continuous and uniformly bounded in $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$.*

PROOF: First note that since $P_\eta \in \mathbf{P_I}$, it follows that $\int vz'\, dP_\eta(x)$ is invertible, while $P_\eta \ll \mu$ and Assumption 3.7(ii) imply $P_\eta(Y_L \le Y_U) = 1$. Therefore, Proposition 2 in Bontemps, Magnac, and Maurin (2012) implies that

$$\text{(E.5)} \qquad \nu\big(p, \Theta_{0,I}(P_\eta)\big) = \int p'\Sigma(P_\eta)^{-1} v\big(y_L + 1\big\{p'\Sigma(P_\eta)^{-1} v > 0\big\}(y_U - y_L)\big)\, dP_\eta(x),$$

provided $P_\eta(Y_L < Y_U) > 0$, while direct calculation shows that (E.5) holds when $P_\eta(Y_L = Y_U) = 1$, since then $\Theta_{0,I}(P_\eta) = \{\Sigma(P_\eta)^{-1} \int v y_L\, dP_\eta(x)\}$. Let $\gamma_\eta(p, v) \equiv p'\Sigma(P_\eta)^{-1} v$ and note that if $(p_n, \eta_n) \to (p_0, \eta_0)$ with $p_0 \in \mathbb{S}^{d_\theta}$, then

$$\text{(E.6)} \qquad \mu\Big((y_L, y_U, v, z): \lim_{n\to\infty} 1\big\{\gamma_{\eta_n}(p_n, v) > 0\big\} = 1\big\{\gamma_{\eta_0}(p_0, v) > 0\big\}\Big) = 1,$$

since $(p, \eta) \mapsto \gamma_\eta(p, v)$ is continuous, and $\mu((y_L, y_U, v, z): p_0'\Sigma(P_{\eta_0})^{-1} v = 0) = 0$ by Assumption 3.7(iii). Moreover,

$$\text{(E.7)} \qquad \lim_{n\to\infty} \sup_{p \in \mathbb{S}^{d_\theta}} \left| \int v^{(i)}(y_U - y_L) 1\big\{\gamma_{\eta_n}(p, v) > 0\big\}\big(h^2_{\eta_n}(x) - h^2_{\eta_0}(x)\big)\, d\mu(x) \right| \le \sup_{x \in \mathcal{X}} 2\|x\|^2 \times \lim_{n\to\infty} \big\{\|h_{\eta_n} - h_{\eta_0}\|_{L^2_\mu} \times \|h_{\eta_n} + h_{\eta_0}\|_{L^2_\mu}\big\} = 0,$$

for any $1 \le i \le d_Z$ by compactness of $\mathcal{X}$, the Cauchy–Schwarz inequality, $\|h_\eta\|_{L^2_\mu} = 1$ for all η, and $\eta \mapsto h_\eta$ being Fréchet differentiable. Hence, compactness of $\mathcal{X}$, result (E.6), and the dominated convergence theorem imply

$$\text{(E.8)} \qquad \lim_{n\to\infty} \int v(y_U - y_L) 1\big\{\gamma_{\eta_n}(p, v) > 0\big\} h^2_{\eta_n}(x)\, d\mu(x) = \int v(y_U - y_L) 1\big\{\gamma_{\eta_0}(p, v) > 0\big\} h^2_{\eta_0}(x)\, d\mu(x).$$

Therefore, for any $p \in \mathbb{S}^{d_\theta}$, we can conclude from (E.8) and $\eta \mapsto \Sigma(P_\eta)^{-1}$ being differentiable by Lemma E.2 that

$$\begin{aligned}
\text{(E.9)}\qquad & \lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int \big(\gamma_{\eta_n}(p,v) - \gamma_{\eta_0}(p,v)\big)(y_U - y_L) \\
&\quad \times 1\big\{\gamma_{\eta_n}(p,v) > 0\big\} h^2_{\eta_n}(x)\, d\mu(x) \\
&= p'\left\{\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\bigg|_{\eta=\eta_0}\right\} \\
&\quad \times \int v(y_U - y_L) 1\big\{\gamma_{\eta_0}(p,v) > 0\big\} h^2_{\eta_0}(x)\, d\mu(x).
\end{aligned}$$

Next, note that $\gamma_{\eta_0}(p,v)(y_U - y_L)$ is uniformly bounded by compactness of $\mathbb{S}^{d_\theta} \times \mathcal{X}$, and hence, arguing as in (A.43),

$$\begin{aligned}
\text{(E.10)}\qquad & \lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int \gamma_{\eta_0}(p,v)(y_U - y_L) 1\big\{\gamma_{\eta_n}(p,v) > 0\big\} \\
&\quad \times \big(h^2_{\eta_n}(x) - h^2_{\eta_0}(x) - 2(\eta_n - \eta_0)\dot{h}_{\eta_0}(x) h_{\eta_0}(x)\big)\, d\mu(x) = 0.
\end{aligned}$$

Thus, results (E.6) and (E.10), compactness of $\mathcal{X}$, and the dominated convergence theorem yield

$$\begin{aligned}
\text{(E.11)}\qquad & \lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int \gamma_{\eta_0}(p,v)(y_U - y_L) \\
&\quad \times 1\big\{\gamma_{\eta_n}(p,v) > 0\big\}\big(h^2_{\eta_n}(x) - h^2_{\eta_0}(x)\big)\, d\mu(x) \\
&= 2\int \gamma_{\eta_0}(p,v)(y_U - y_L) 1\big\{\gamma_{\eta_0}(p,v) > 0\big\} \dot{h}_{\eta_0}(x) h_{\eta_0}(x)\, d\mu(x).
\end{aligned}$$

In addition, Lemma E.2 and the mean value theorem imply that, for some $\bar{\eta}_n(x)$ between η_n and η_0,

$$\begin{aligned}
\text{(E.12)}\qquad & \lim_{n\to\infty} \bigg| \int \gamma_{\eta_0}(p,v)(y_U - y_L) \\
&\quad \times \big(1\big\{\gamma_{\eta_n}(p,v) > 0\big\} - 1\big\{\gamma_{\eta_0}(p,v) > 0\big\}\big) h^2_{\eta_0}(x)\, d\mu(x)\bigg| \\
&= \lim_{n\to\infty} \bigg| \int \gamma_{\eta_0}(p,v)(y_U - y_L) \\
&\quad \times \left(1\left\{\gamma_{\eta_0}(p,v) > (\eta_0 - \eta_n)\frac{\partial}{\partial\eta}\gamma_\eta(p,v)\bigg|_{\eta=\bar{\eta}_n(x)}\right\}\right.
\end{aligned}$$

$$- 1\{\gamma_{\eta_0}(p, v) > 0\}\Big) h^2_{\eta_0}(x)\, d\mu(x)\Big|$$
$$\leq \lim_{n\to\infty} \int \big|\gamma_{\eta_0}(p, v)(y_U - y_L)\big|$$
$$\times 1\{\big|\gamma_{\eta_0}(p, v)\big| \leq M|\eta_0 - \eta_n|\} h^2_{\eta_0}(x)\, d\mu(x),$$

where the inequality holds for some $M > 0$ due to Lemma E.2 and compactness of $\mathbb{S}^{d_\theta} \times \mathcal{X}$ implying $\frac{\partial}{\partial\eta}\gamma_\eta(p, v)|_{\eta=\eta_0}$ is uniformly bounded for η_0 in a neighborhood of zero. Therefore, from (E.12) we conclude

$$\text{(E.13)} \quad \lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|}\bigg|\int \gamma_{\eta_0}(p, v)(y_U - y_L)$$
$$\times \big(1\{\gamma_{\eta_n}(p, v) > 0\} - 1\{\gamma_{\eta_0}(p, v) > 0\}\big) h^2_{\eta_0}(x)\, d\mu(x)\bigg|$$
$$\leq 2 \sup_{x\in\mathcal{X}} \|x\| \times \lim_{n\to\infty} M \int 1\{\big|\gamma_{\eta_0}(p, v)\big| \leq M|\eta_0 - \eta_n|\} h^2_{\eta_0}(x)\, d\mu(x)$$
$$= 0,$$

where the final equality results from the monotone convergence theorem, and $\mu((y_L, y_U, v, z) : p'\Sigma(P_{\eta_0})^{-1}v = 0) = 0$ by Assumption 3.7(iii) and $p'\Sigma(P_{\eta_0})^{-1} \neq 0$. Finally, combining results (E.9), (E.11), and (E.13), we can obtain

$$\text{(E.14)} \quad \frac{\partial}{\partial\eta}\bigg\{\int \gamma_\eta(p, v)(y_U - y_L) 1\{\gamma_\eta(p, v) > 0\} h^2_\eta(x)\, d\mu(x)\bigg\}\bigg|_{\eta=\eta_0}$$
$$= \int \bigg(p'\bigg\{\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\bigg|_{\eta=\eta_0}\bigg\} v h^2_{\eta_0}(x) + 2\gamma_{\eta_0}(p, v)\dot{h}_{\eta_0}(x) h_{\eta_0}(x)\bigg)$$
$$\times (y_U - y_L) 1\{\gamma_{\eta_0}(p, v) > 0\}\, d\mu(x).$$

Similarly, Lemma E.2, compactness of $\mathcal{X}$, and arguing as in (E.9) and (E.11) allow us to establish that

$$\text{(E.15)} \quad \frac{\partial}{\partial\eta}\bigg\{\int \gamma_\eta(p, v) y_L h^2_\eta(x)\, d\mu(x)\bigg\}\bigg|_{\eta=\eta_0}$$
$$= \int \bigg(p'\bigg\{\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\bigg|_{\eta=\eta_0}\bigg\} v h^2_{\eta_0}(x)$$
$$+ 2\gamma_{\eta_0}(p, v)\dot{h}_{\eta_0}(x) h_{\eta_0}(x)\bigg) y_L\, d\mu(x).$$

Result (E.4) then follows from (E.14), (E.15), Lemma E.2, and the definitions of ψ_ν and ψ_Σ.

To establish continuity, let $(p_n, \eta_n) \to (p_0, \eta_0) \in \mathbb{S}^{d_\theta} \times N$. Results (E.6) and (E.7) then imply that

$$\begin{aligned}
\text{(E.16)} \qquad & \lim_{n\to\infty} \int v(y_U - y_L) 1\{\gamma_{\eta_n}(p_n, v) > 0\} h^2_{\eta_n}(x)\, d\mu(x) \\
&= \int v(y_U - y_L) 1\{\gamma_{\eta_0}(p_0, v) > 0\} h^2_{\eta_0}(x)\, d\mu(x)
\end{aligned}$$

by the dominated convergence theorem. Next, note that by compactness of $\mathcal{X}$ and the Cauchy–Schwarz inequality,

$$\begin{aligned}
\text{(E.17)} \qquad & \lim_{n\to\infty} \bigg| \int v^{(i)}(y_U - y_L) \\
&\quad \times 1\{\gamma_{\eta_n}(p_n, v) > 0\} \big(\dot{h}_{\eta_n}(x) h_{\eta_n}(x) - \dot{h}_{\eta_0}(x) h_{\eta_0}(x)\big)\, d\mu(x) \bigg| \\
&\le 2 \sup_{x \in \mathcal{X}} \|x\|^2 \\
&\quad \times \lim_{n\to\infty} \{ \|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L^2_\mu} \|h_{\eta_n}\|_{L^2_\mu} + \|h_{\eta_n} - h_{\eta_0}\|_{L^2_\mu} \|\dot{h}_{\eta_0}\|_{L^2_\mu} \} \\
&= 0,
\end{aligned}$$

since $\|h_\eta\|_{L^2_\mu} = 1$ for all η and $\eta \mapsto h_\eta$ is continuously Fréchet differentiable. Hence, we can conclude that

$$\begin{aligned}
\text{(E.18)} \qquad & \lim_{n\to\infty} \int v(y_U - y_L) 1\{\gamma_{\eta_n}(p_n, v) > 0\} \dot{h}_{\eta_n}(x) h_{\eta_n}(x)\, d\mu(x) \\
&= \int v(y_U - y_L) 1\{\gamma_{\eta_0}(p_0, v) > 0\} \dot{h}_{\eta_0}(x) h_{\eta_0}(x)\, d\mu(x)
\end{aligned}$$

by (E.6) and the dominated convergence theorem. Therefore, (E.14), (E.16), (E.18), and Lemma E.2 yield

$$\begin{aligned}
\text{(E.19)} \qquad & \lim_{n\to\infty} \frac{\partial}{\partial\eta} \left\{ \int \gamma_\eta(p_n, v)(y_U - y_L) 1\{\gamma_\eta(p_n, v) > 0\} h^2_\eta(x)\, d\mu(x) \right\} \bigg|_{\eta=\eta_n} \\
&= \frac{\partial}{\partial\eta} \left\{ \int \gamma_\eta(p_0, v)(y_U - y_L) 1\{\gamma_\eta(p_0, v) > 0\} h^2_\eta(x)\, d\mu(x) \right\} \bigg|_{\eta=\eta_0}.
\end{aligned}$$

Similarly, employing the same arguments as in (E.16) and (E.18) together with result (E.15), it is possible to show:

$$\text{(E.20)} \quad \lim_{n\to\infty} \frac{\partial}{\partial\eta}\left\{\int \gamma_\eta(p_n, v) y_L h_\eta^2(x)\, d\mu(x)\right\}\bigg|_{\eta=\eta_n} = \frac{\partial}{\partial\eta}\left\{\int \gamma_\eta(p_0, v) y_L h_\eta^2(x)\, d\mu(x)\right\}\bigg|_{\eta=\eta_0}.$$

Thus, continuity of $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ follows from (E.5), (E.19), and (E.20). Finally, note that since $\eta \mapsto h_\eta$ is continuously Fréchet differentiable, we may choose the neighborhood $N \subseteq \mathbf{R}$ so that $\|\dot{h}_\eta\|_{L^2_\mu}$ is uniformly bounded in $\eta \in N$. The Cauchy–Schwarz inequality then implies $|\int \dot{h}_\eta(x) h_\eta(x)\, d\mu(x)| \leq \|\dot{h}_\eta\|_{L^2_\mu}\|h_\eta\|_{L^2_\mu} < \infty$ uniformly in $\eta \in N$. Therefore, compactness of $\mathcal{X} \times \mathbb{S}^{d_\theta}$, Lemma E.2, and results (E.5), (E.14), and (E.15) imply $\frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ is uniformly bounded in $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$, and the lemma follows. *Q.E.D.*

LEMMA E.4: *Let Assumption* 3.7 *hold, and* $\rho_I : \mathbf{P_I} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *be given by* $\rho_I(P) \equiv \nu(\cdot, \Theta_{0,I}(P))$. *Then* ρ_I *is pathwise weak-differentiable at any* $P \in \mathbf{P_I}$, *and for* $s \equiv \sqrt{dP/d\mu}$, *the derivative* $\dot{\rho}_I : \dot{\mathbf{S}}_\mathbf{I} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *satisfies*

$$\dot{\rho}_I(\dot{h}_0)(p) = 2\int \{\psi_\nu(p, x, P) - \psi_\Sigma(p, x, P)\}\dot{h}_0(x) h_0(x)\, d\mu(x),$$

where ψ_ν *and* ψ_Σ *are as defined in equations* (16) *and* (17), *respectively*.

PROOF: We first note that Lemma E.3 implies $\dot{\rho}_I(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ for any $\dot{h}_0 \in \dot{\mathbf{S}}_\mathbf{I}$. In addition, $\dot{\rho}_I$ is linear by inspection, while $\psi_\nu(p, x, P)$ and $\psi_\Sigma(p, x, P)$ being uniformly bounded in $(p, x) \in \mathbb{S}^{d_\theta} \times \mathcal{X}$ by Assumption 3.7(i) imply

$$\text{(E.21)} \quad \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \|\dot{\rho}_I(\dot{h}_0)\|_\infty \leq \sup_{(p,x)\in\mathbb{S}^{d_\theta}\times\mathcal{X}} 2\{|\psi_\nu(p, x, P)| + |\psi_\Sigma(p, x, P)|\} \times \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \{\|\dot{h}_0\|_{L^2_\mu} \times \|h_0\|_{L^2_\mu}\} < \infty,$$

and hence $\dot{\rho}_I$ is continuous as well. Moreover, for any finite Borel measure B on $\mathbb{S}^{d_\theta}$ and curve $\eta \mapsto P_\eta \in \mathbf{P_I}$ with $h_0 = s$, the mean value and dominated convergence theorems together with Lemma E.3 yield

$$\text{(E.22)} \quad \lim_{\eta_0\to 0} \int \left\{\frac{\nu(p, \Theta_{0,I}(P_{\eta_0})) - \nu(p, \Theta_{0,I}(P))}{\eta_0} - \dot{\rho}_I(\dot{h}_0)(p)\right\} dB(p) = 0$$

(see also (A.57)). Result (E.22) verifies that $\dot{\rho}_I$ is the weak derivative of ρ_I, and the lemma follows. *Q.E.D.*

PROOF OF THEOREM 3.3: As in the proof of Theorem 3.2, we let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{S}^{d_\theta}$, which is the dual of $\mathbf{B}$ by Corollary 14.15 in Aliprantis and Border (2006). Let $\rho_I : \mathbf{P_I} \to \mathbf{B}$ be given by $\rho_I(P) \equiv \nu(\cdot, \Theta_{0,I}(P))$, which has weak derivative $\dot{\rho}_I$ by Lemma E.4. For any $B \in \mathbf{B}^*$, then define

$$\dot{\rho}_I^T(B)(x) \equiv 2\int_{\mathbb{S}^{d_\theta}} \{\psi(x,p,P) - E[\psi(X_i,p,P)]\} s(x)\,dB(p), \tag{E.23}$$

where $s \equiv \sqrt{dP/d\mu}$, and the measurability of the integrand can be established arguing as in (A.59). In what follows, we aim to show $\dot{\rho}_I^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{I}$ is the adjoint of $\dot{\rho}_I : \dot{\mathbf{S}}_\mathbf{I} \to \mathbf{B}$. To this end, note that $\dot{\rho}_I^T(B) \in L^2_\mu$ for any $B \in \mathbf{B}^*$ since $\psi(p,x,P) = \psi_\nu(p,x,P) - \psi_\Sigma(p,x,P)$ is uniformly bounded in $(p,x) \in \mathbb{S}^{d_\theta} \times \mathcal{X}$, as argued in (E.21). Moreover,

$$\begin{aligned}\int_{\mathcal{X}} \dot{\rho}_I^T(B)(x) s(x)\,d\mu(x) \\ = 2\int_{\mathbb{S}^{d_\theta}} \int_{\mathcal{X}} \{\psi(x,p,P) - E[\psi(X_i,p,P)]\}\,dP(x)\,dB(p) = 0,\end{aligned} \tag{E.24}$$

by exchanging the order of integration and exploiting that $s^2 = dP/d\mu$. Hence, Lemma E.1 and (E.24) verify that $\dot{\rho}_I^T(B) \in \dot{\mathbf{S}}_\mathbf{I}$ for any $B \in \mathbf{B}^*$. Finally, for any $\dot{h}_0 \in \dot{\mathbf{S}}_\mathbf{I}$ and $B \in \mathbf{B}^*$, we can use that $\int \dot{h}_0(x)s(x)\,d\mu(x) = 0$ by Lemma E.1, exchange the order of integration, and exploit Lemma E.4 to obtain that

$$\begin{aligned}\int_{\mathcal{X}} \dot{\rho}_I^T(B)(x)\dot{h}_0(x)\,d\mu(x) &= \int_{\mathbb{S}^{d_\theta}} \int_{\mathcal{X}} \psi(x,p,P)\dot{h}_0(x)s(x)\,d\mu(x)\,dB(p) \\ &= \int_{\mathbb{S}^{d_\theta}} \dot{\rho}_I(\dot{h}_0)(p)\,dB(p).\end{aligned} \tag{E.25}$$

From result (E.25), we conclude that $\dot{\rho}_I^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{I}$ is indeed the adjoint of $\dot{\rho}_I : \dot{\mathbf{S}}_\mathbf{I} \to \mathbf{B}$, and the theorem then follows from Theorem 5.2.1 in Bickel et al. (1993). *Q.E.D.*

The principal challenge in establishing Theorem 3.3 is in verifying pathwise weak-differentiability of the support function of the identified set. Differentiability of the support function in particular implies that the scalar valued parameter $Q \mapsto \nu(p_0, \Theta_{0,I}(Q))$ must be differentiable at every $p_0 \in \mathbb{S}^{d_\theta}$, which, by (15), is equivalent to

$$\text{(E.26)}\quad \nu\big(p_0, \Theta_{0,I}(P_\eta)\big) = \int p_0'\Sigma(P_\eta)^{-1}v\big(y_L + 1\big\{p_0'\Sigma(P_\eta)^{-1}v > 0\big\}(y_U - y_L)\big)\,dP_\eta(x)$$

being differentiable in η for any parametric submodel $\eta \mapsto P_\eta$. Inspecting (E.26), however, reveals that nondifferentiability at $\eta = 0$ may occur if $P(p_0'\Sigma(P)^{-1}V = 0) > 0$—a situation that is ruled out by Assumption 3.7(iii). Interestingly, when V is a discrete random vector, the identified set $\Theta_{0,I}(P)$ has "flat" or "exposed" faces, and the $p_0 \in \mathbb{S}^{d_\theta}$ such that $P(p_0'\Sigma(P)^{-1}V = 0) > 0$ are precisely the $p_0 \in \mathbb{S}^{d_\theta}$ that are orthogonal to these flat faces; see Bontemps, Magnac, and Maurin (2012). In close connection to Remark 3.2, it is then possible to show that $Q \mapsto \nu(p_0, \Theta_{0,I}(Q))$ is not pathwise weak-differentiable at any such p_0 by constructing a path $\eta \mapsto P_\eta$ that alters the slope of the exposed face.

EXAMPLE E.1: Suppose $Z = V = (1, W)'$, $W \in \{-1, 0, 1\}$, and $Y_L, Y_U \in \mathcal{Y} \subset \mathbf{R}$ with $\mathcal{Y}$ compact. Further let $X = (Y_L, Y_U, V')'$, $\mathcal{X} = \mathcal{Y} \times \mathcal{Y} \times \{1\} \times \{-1, 0, 1\}$, and $\mu \in \mathbf{M}$ satisfy Assumption 3.7(ii). The set of $\theta = (\alpha, \beta)'$ with

$$\text{(E.27)}\quad E[\tilde{Y} - \alpha - W\beta] = 0, \quad E\big[W(\tilde{Y} - \alpha - W\beta)\big] = 0,$$

for some $\tilde{Y}$ satisfying $Y_L \leq \tilde{Y} \leq Y_U$, then constitutes the identified set under P. Further suppose P is such that

$$\text{(E.28)}\quad P(W = -1) = P(W = 0) = P(W = 1) = \frac{1}{3},$$

for $a \in \{-1, 0, 1\}$ and $\ell \in \{L, U\}$ define $E_P[Y_\ell | W = a] \equiv \int y_\ell 1\{w = a\}\,dP(x)/P(W = a)$, and for simplicity let

$$\text{(E.29)}\quad E_P[Y_L | W = 0] = E_P[Y_U | W = 0] = 0.$$

Let us consider a submodel satisfying $E_{P_\eta}[Y_\ell | W = a] = E_P[Y_\ell | W = a]$ for all $a \in \{-1, 0, 1\}$ and $\ell \in \{L, U\}$, and

$$\text{(E.30)}\quad \begin{aligned} P_\eta(W = -1) &= \frac{1}{3}(1 - \eta), \\ P_\eta(W = 0) &= \frac{1}{3}(1 + 2\eta), \\ P_\eta(W = 1) &= \frac{1}{3}(1 - \eta). \end{aligned}$$

Along the submodel $\eta \mapsto P_\eta$, we can then obtain by direct calculation that the identified set at P_η is given by

$$\begin{aligned}\text{(E.31)}\quad & \Theta_0(P_\eta) \\ & = \Big\{\theta \in \mathbf{R}^2 : \text{(i) } E_{P_\eta}[Y_L|W=-1] \le \frac{3}{2}\frac{\alpha}{1-\eta} - \beta \le E_{P_\eta}[Y_U|W=-1], \\ & \quad \text{(ii) } E_{P_\eta}[Y_L|W=1] \le \frac{3}{2}\frac{\alpha}{1-\eta} + \beta \le E_{P_\eta}[Y_U|W=1]\Big\}.\end{aligned}$$

Thus, $\Theta_0(P_\eta)$ is a parallelogram with the slope of exposed faces depending on η. As in Remark 3.2, $\eta \mapsto \nu(p_0, \Theta_0(P_\eta))$ is not differentiable at $\eta = 0$ for an appropriate choice of p_0. For instance, for $p_0 = (\frac{3}{\sqrt{13}}, \frac{2}{\sqrt{13}})$, we obtain by (E.26)

$$\begin{aligned}\text{(E.32)}\quad \nu\big(p_0, \Theta_0(P_\eta)\big) = {} & \frac{2-\eta}{\sqrt{13}} E[Y_U|W=1] - \frac{\eta}{\sqrt{13}}\big(E[Y_L|W=-1] \\ & + E[Y_U - Y_L|W=-1]1\{\eta < 0\}\big),\end{aligned}$$

which is not differentiable at $\eta = 0$ if $E[Y_U - Y_L|W = -1] \neq 0$. Thus, $\eta \mapsto \nu(p_0, \Theta_0(P_\eta))$ is not differentiable at $\eta = 0$ precisely at a p_0 that is orthogonal to one of the exposed faces of the identified set $\Theta_0(P)$.

APPENDIX F: DISCUSSION OF EXAMPLES 2.1, 2.2, 2.3, AND 2.4

In this appendix, we revisit Examples 2.1, 2.2, 2.3, and 2.4 from the main text. We map each example into our general framework, and examine Assumptions 3.2, 3.3, 3.4, 3.5, and 3.6 in their context.

EXAMPLE 2.1—Interval Censored Outcome:
In this example, $X = (Y_L, Y_U, Z')'$ and we let $\mathcal{Y} \subseteq \mathbf{R}$, $\mathcal{Z} = \{z_1, \dots, z_K\}$ with $K < \infty$, and $\mathcal{X} = \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z}$. For $\mathbf{M}$ the set of Borel probability measures on $\mathcal{X}$ and any $Q \in \mathbf{M}$ such that $Q(Z = z_k) > 0$, then denote, for $\ell \in \{L, U\}$,

$$\text{(F.1)}\qquad E_Q[Y_\ell|Z = z_k] \equiv \frac{\int y_\ell 1\{Z = z_k\}\, dQ(x)}{\int 1\{Z = z_k\}\, dQ(x)}.$$

For a parameter space $\Theta \subseteq \mathbf{R}^{d_\theta}$ and any $Q \in \mathbf{M}$, then recall that, in this example, the identified set under Q is

$$\begin{aligned}\text{(F.2)}\qquad \Theta_0(Q) \equiv \big\{&\theta \in \Theta : E_Q[Y_L|Z = z_k] \le z_k'\theta \le E_Q[Y_U|Z = z_k] \\ &\text{for all } 1 \le k \le K\big\}.\end{aligned}$$

To map this setting into the framework of (2) and (3), we let $1_{\mathcal{Z}}(z) \equiv (1\{z = z_1\}, \dots, 1\{z = z_K\})'$ and $m_S(x, \theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ for all $\theta \in \Theta$. Then define $F_S : \mathbf{R}^{3K} \to \mathbf{R}^{2K}$ to be pointwise given by

$$F_S^{(i)}(v) \equiv \begin{cases} \dfrac{v^{(i)}}{v^{(2K+i)}}, & i = 1, \dots, K, \\ -\dfrac{v^{(i)}}{v^{(K+i)}}, & i = K+1, \dots, 2K. \end{cases} \tag{F.3}$$

If $Q \in \mathbf{M}$ satisfies $Q(Z = z_k) > 0$ for some $1 \leq k \leq K$, then (F.3) implies $F_S^{(k)}(\int m_S(x, \theta)\, dQ(x)) = E_Q[Y_L | Z = z_k]$ and $F_S^{(2k)}(\int m_S(x, \theta)\, dQ(x)) = -E_Q[Y_U | Z = z_k]$. Hence, setting $A = (-z_1, \dots, -z_K, z_1, \dots, z_K)'$, we obtain

$$\Theta_0(Q) = \left\{\theta \in \Theta : A\theta + F_S\left(\int m_S(x, \theta)\, dQ(x)\right) \leq 0\right\}. \tag{F.4}$$

The following more primitive assumptions suffice for verifying Assumptions 3.2–3.6 in this example.

ASSUMPTION F.1: (i) $\mathcal{Y}$ is compact; (ii) $\Theta \equiv \{\theta \in \mathbf{R}^{d_\theta} : \|\theta\|^2 \leq B_0\}$ with $B_0 < \infty$ satisfying $C_0 B_0 > K\{\sup_{y \in \mathcal{Y}} y^2\}$, where $C_0 \equiv \inf_{p \in \mathbb{S}^{d_\theta}} \sum_k \langle p, z_k \rangle^2$; (iii) $K \geq d_\theta$; (iv) any subset $\mathcal{C} \subseteq \mathcal{Z}$ with $\#\mathcal{C} \leq d_\theta$ is linearly independent.

ASSUMPTION F.2: (i) For some $\theta_0 \in \mathbf{R}^{d_\theta}$, $E_P[Y_L | Z = z_k] \leq z_k' \theta_0 \leq E_P[Y_U | Z = z_k]$ for all $1 \leq k \leq K$; (ii) $P(Z = z_k) > 0$ and $E_P[Y_L - Y_U | Z = z_k] < 0$ for all $1 \leq k \leq K$; (iii) $\#\mathcal{A}(\theta, P) \leq d_\theta$ for all $\theta \in \Theta_0(P)$.

Assumption F.1(i) imposes that Y_L and Y_U have compact support, which we require to verify Assumption 3.4(i). Assumption F.1(ii) defines Θ to be a ball of radius $\sqrt{B_0}$, where B_0 is chosen to ensure that $\Theta_0(P) \subset \Theta^o$ as required by Assumption 3.6(i). Assumption F.1(iii)–(iv) imposes a linear independence restriction on the support points of Z, which together guarantee that $\Theta_0(P)$ is bounded. Assumption F.2 contains the main requirements on P. In particular, Assumption F.2(i), which holds if the model is properly specified, guarantees that $\Theta_0(P) \neq \emptyset$. The requirement $E_P[Y_L - Y_U | Z = z_k] < 0$ ensures that there is no $\theta \in \Theta_0(P)$ such that $E_P[Y_L | Z = z_k] = z_k' \theta = E_P[Y_U | Z = z_k]$, which would violate Assumption 3.6(iv). Finally, Assumption F.2(iii) requires that the number of binding constraints at each $\theta \in \Theta_0(P)$ be less than or equal to d_θ, and, together with Assumption F.1(iv), implies Assumption 3.6(iv). We note that if $K = d_\theta$, then Assumption F.1(iv) and $E_P[Y_L - Y_U | Z = z_k] < 0$ imply that Assumption F.2(iii) is automatically satisfied. In general, however, Assumption F.2(iii) imposes additional requirements on P.

PROPOSITION F.1: *In Example* 2.1, *Assumptions* F.1 *and* F.2 *imply Assumptions* 3.2–3.6.

PROOF: Assumption 3.2 is implied by Assumption F.1(ii). Further note that since the $2K \times d_\theta$ matrix A is known, Assumption 3.3 holds. Moreover, since $\mathcal{Y}$ is compact by Assumption F.1(i), $m_S(x,\theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ is uniformly bounded in $\mathcal{X} \times \Theta$ and hence $m(x,\theta) = (m_S(x,\theta)', \theta' A')'$ and Θ being compact by Assumption F.1(ii) verify Assumption 3.4(i). In addition, given the definition of $m_S(x,\theta)$, Assumption 3.4(ii)–(iii) directly follows from

$$\nabla_\theta m(x,\theta) = \nabla_\theta \begin{bmatrix} m_S(x,\theta) \\ A\theta \end{bmatrix} = \begin{bmatrix} 0 \\ A \end{bmatrix}. \tag{F.5}$$

In order to verify Assumption 3.5, set $0 < \varepsilon_0 < \inf_k P(Z = z_k)$, which is possible by Assumption F.2(ii), and $M_0 > 0$ so that $\max\{\sup_{y\in\mathcal{Y}} |y|, B_0 \sup_{z\in\mathcal{Z}} \|z\|\} < M_0 < \infty$, which is possible by compactness of $\mathcal{Y}$. Then defining

$$V_0 \equiv (-M_0, M_0)^{2K} \times (\varepsilon_0, 1)^K \times (-M_0, M_0)^{2K}, \tag{F.6}$$

and noting that $F(v)$ is differentiable unless $v^{(i)} = 0$ for some $2K+1 \le i \le 3K$, it follows that Assumption 3.5(i) holds. Moreover, since ∇F is continuous on the closure of V_0 and V_0 is precompact, Assumption 3.5(ii) holds as well.

We next verify that P satisfies Assumption 3.6. First observe that Assumption F.2(i) implies $\theta_0 \in \Theta_0(P)$ and hence $\Theta_0(P) \ne \emptyset$. Next, also note that if $\theta \in \Theta_0(P)$, then (F.2) implies that for, any $1 \le k \le K$,

$$|z_k'\theta| \le \max\{|E_P[Y_L|Z = z_k]|, |E_P[Y_U|Z = z_k]|\} \le \sup_{y\in\mathcal{Y}} |y|. \tag{F.7}$$

Furthermore, Assumption F.1(iii)–(iv) implies $\mathbf{R}^{d_\theta} = \operatorname{span}\{z_1, \dots, z_K\}$, and hence $C_0 = \inf_{p\in\mathbb{S}^{d_\theta}} \sum_k \langle p, z_k\rangle^2 > 0$ by compactness of $\mathbb{S}^{d_\theta}$. Therefore, since $\theta/\|\theta\| \in \mathbb{S}^{d_\theta}$, we obtain from (F.7) that, for any $\theta \in \Theta_0(P)$,

$$\|\theta\|^2 C_0 \le \|\theta\|^2 \sum_{k=1}^K \left\langle z_k, \frac{\theta}{\|\theta\|}\right\rangle^2 \le K \sup_{y\in\mathcal{Y}} y^2. \tag{F.8}$$

It then follows from Assumption F.1(ii) that if $\theta \in \Theta_0(P)$, then $\|\theta\|^2 < B_0$ and hence $\Theta_0(P) \subseteq \Theta^o$. However, since $\Theta_0(P)$ is closed, we must have $\Theta_0(P) \subset \Theta^o$, which verifies Assumption 3.6(i).

Since $m_S(x,\theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ does not depend on θ, it follows that $\mathcal{S}_i = \emptyset$ for all $1 \le i \le 2K$ (see (4)), and hence Assumption 3.6(ii) actually holds for all $Q \in \mathbf{M}$. In turn, by definitions of ε_0 and M_0, we also have $\int m(x,\theta)\,dP(x) \in V_0$ for all $\theta \in \Theta$ and thus Assumption 3.6(iii) holds as well. Finally, note that

$$\begin{aligned} &\nabla F^{(i)}\left(\int m(x,\theta)\,dP(x)\right) \int \nabla_\theta m(x,\theta)\,dP(x) \\ &\quad = \begin{cases} -z_i, & \text{if } 1 \le i \le K, \\ +z_i, & \text{if } K+1 \le i \le 2K. \end{cases} \end{aligned} \tag{F.9}$$

For notational simplicity, let $\mathcal{P}(\theta) = \{\nabla F^{(i)}(\int m(x,\theta)\,dP(x))\int \nabla_\theta m(x,\theta)\,dP(x)\}_{i\in\mathcal{A}(\theta,P)}$. Then note that since $E_P[Y_L - Y_U|Z = z_k] < 0$, it follows, for $1 \le i \le K$, that if $i \in \mathcal{A}(\theta, P)$, then $K + i \notin \mathcal{A}(\theta, P)$—or equivalently, if $-z_i \in \mathcal{P}(\theta)$, then $z_i \notin \mathcal{P}(\theta)$. Assumptions F.1(iv) and F.2(iii) then imply that the elements of $\mathcal{P}(\theta)$ are linearly independent for all $\theta \in \Theta_0(P)$, which verifies Assumption 3.6(iv). *Q.E.D.*

EXAMPLE 2.2—Discrete Choice:

The structure of this example is identical to that of Example 2.1, though the notation is substantially more cumbersome. In this example, $X = (Y', Z^{*\prime})'$, and we let $\mathcal{Y} \subseteq \mathbf{R}^{d_Y}$. Also recall Z^* is assumed to have finite support $\mathcal{Z} = \{z_1, \dots, z_K\}$ with $K < \infty$. Set $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, and let $\mathbf{M}$ denote the set of Borel measures on $\mathcal{X}$. For notational convenience, we also define $\Delta(y, z_j, z_k) \equiv \psi(y, z_j) - \psi(y, z_k)$ and the set $\mathcal{V}$ to be given by

$$\mathcal{V} \equiv \{z_1 - z_2, \dots, z_1 - z_K, z_2 - z_3, \dots, z_2 - z_K, \dots, z_{K-1} - z_K\}. \tag{F.10}$$

For any $Q \in \mathbf{M}$ such that $Q(Z^* = z_k) > 0$, let $E_Q[\Delta(Y, z_j, z_k)|Z^* = z_k] \equiv \int \Delta(y, z_j, z_k)1\{z^* = z_k\}\,dQ(x)/\int 1\{z^* = z_k\}\,dQ(x)$ (as in (F.1)), and note that for a parameter space Θ, the identified set under $Q \in \mathbf{M}$ in this example is

$$\Theta_0(Q) \equiv \big\{\theta \in \Theta : E_Q\big[\Delta(Y, z_j, z_k)|Z^* = z_k\big] + (z_j - z_k)'\theta \le 0 \text{ for all } z_j \ne z_k\big\}. \tag{F.11}$$

To identify (F.11) with the framework of (2) and (3), for each $1 \le k \le K$, let $v_k(y, z^*) \in \mathbf{R}^{K-1}$ satisfy

$$v_k^{(j)}\big(y, z^*\big) = \begin{cases} \Delta(y, z_j, z_k)1\big\{z^* = z_k\big\}, & 1 \le j < k, \\ \Delta(y, z_{j+1}, z_k)1\big\{z^* = z_k\big\}, & k \le j \le K-1. \end{cases} \tag{F.12}$$

Then let $v(y, z^*) = (v_1(y, z^*)', \dots, v_K(y, z^*)')'$, $1_{\mathcal{Z}}(z^*) \equiv (1\{z^* = z_1\}, \dots, 1\{z^* = z_K\})'$ and set $m_S(x, \theta) \in \mathbf{R}^{K^2}$ to be given by $m_S(x, \theta) = (v(y, z^*)', 1_{\mathcal{Z}}(z^*)')'$. We can then define $F_S : \mathbf{R}^{K^2} \to \mathbf{R}^{K(K-1)}$ to be pointwise given by

$$F_S^{(i)}(v) = \frac{v^{(i)}}{v^{(K(K-1)+\lceil i/(K-1)\rceil)}}, \quad i = 1, \dots, K(K-1), \tag{F.13}$$

where $\lceil c \rceil$ denotes the smallest integer k such that $k \ge c$. Given these definitions, if $Q \in \mathbf{M}$ is such that $Q(Z^* = z_k) > 0$ and $(K-1)(k-1) + 1 \le i \le (K-1)k$, then $F_S^{(i)}(\int m_S(x, \theta)\,dQ(x)) = E_Q[\Delta(Y, z_j, z_k)|Z^* = z_k]$ for some $j \ne k$. Moreover, by setting $A = ((z_1 - z_2), \dots, (z_1 - z_K), \dots, (z_K - z_1), \dots, (z_K - z_{K-1}))'$, we obtain

$$\Theta_0(Q) = \left\{\theta \in \Theta : A\theta + F_S\left(\int m_S(x, \theta)\,dQ(x)\right) \le 0\right\}. \tag{F.14}$$

Given the identical structure of Examples 2.1 and 2.2, we can derive sufficient conditions for Assumptions 3.2–3.6 by recasting Assumptions F.1 and F.2 in the present context. A formal proof that Assumptions F.3 and F.4 imply Assumptions 3.2–3.6 can be obtained by arguments identical to those of Proposition F.1 and is therefore omitted.

ASSUMPTION F.3: (i) $\psi : \mathcal{Y} \times \mathcal{Z} \to \mathbf{R}$ is bounded; (ii) $\Theta \equiv \{\theta \in \mathbf{R}^{d_\theta} : \|\theta\|^2 \leq B_0\}$ with $B_0 < \infty$ satisfying $2C_0B_0 > K(K-1)\{\sup_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}(\psi(y,z))^2\}$, where $C_0 \equiv \inf_{p\in\mathbb{S}^{d_\theta}} \sum_{v\in\mathcal{V}} \langle p, v\rangle^2$; (iii) $K(K-1) \geq 2d_\theta$; (iv) any subset $\mathcal{C} \subseteq \mathcal{V}$ satisfying $\#\mathcal{C} \leq d_\theta$ is linearly independent.

ASSUMPTION F.4: (i) For some $\theta_0 \in \mathbf{R}^{d_\theta}$, $E_P[\Delta(Y, z_j, z_k)|Z^* = z_k] + (z_j - z_k)'\theta_0 \leq 0$ for all $z_j \neq z_k \in \mathcal{Z}$; (ii) $P(Z^* = z_k) > 0$ for all $1 \leq k \leq K$; (iii) $E_P[\Delta(Y, z_j, z_k)|Z^* = z_j] \neq E_P[\Delta(Y, z_j, z_k)|Z^* = z_k]$ for any $1 \leq j < k \leq K$; (iv) $\#\mathcal{A}(\theta, P) \leq d_\theta$ for all $\theta \in \Theta_0(P)$.

Assumption F.3(i) guarantees $m(x, \theta)$ is bounded as required by Assumption 3.4(i). As in Assumption F.1(ii), $\Theta \subset \mathbf{R}^{d_\theta}$ is defined to be a sufficiently large sphere to ensure that $\Theta_0(P) \subset \Theta^o$, as demanded by Assumption 3.6(i). The gradient $\nabla F^{(i)}(\int m(x,\theta)\,dP(x)) \int \nabla_\theta m(x,\theta)\,dP(x)$ at each active constraint is of the form $(z_j - z_k)$ for some $z_j \neq z_k \in \mathcal{Z}$. Therefore, to ensure that Assumption 3.6(iv) holds, we must rule out that a $\theta \in \Theta_0(P)$ satisfies

$$\text{(F.15)} \quad \begin{aligned} &E\big[\Delta(Y, z_j, z_k)|Z^* = z_k\big] + (z_j - z_k)'\theta \\ &\quad = 0 = E\big[\Delta(Y, z_k, z_j)|Z^* = z_j\big] + (z_k - z_j)'\theta, \end{aligned}$$

which is guaranteed by Assumption F.4(iii). A consequence of Assumption F.4(iii) is that when Z^* has K points of support, it generates $K(K-1)$ constraints, of which at most $K(K-1)/2$ can be active. For this reason, Assumption F.3(iii) requires $K(K-1)/2 \geq d_\theta$, which, together with Assumption F.3(iv), implies $\Theta_0(P)$ is bounded. Assumption F.4(i) is satisfied if the model is properly specified and implies $\Theta_0(P) \neq \emptyset$. Finally, Assumptions F.3(iv) and F.4(iv) together provide a sufficient condition for Assumption 3.6(iv) to be satisfied.

REMARK F.1: The moment inequalities in (6) are a special case of a larger system implied by the optimality condition in (5). In particular, for any $\mathcal{F}$ measurable random variable V, equation (5) implies that, for any $z_j \in \mathcal{Z}$,

$$\text{(F.16)} \quad E\big[\big(\big(\psi(Y, z_j) - \psi\big(Y, Z^*\big)\big) + \big(z_j - Z^*\big)'\theta\big)g(V)\big] \leq 0,$$

provided $g(V) \geq 0$ almost surely; see, for example, Ho (2009). Indeed, note that (F.16) reduces to (6) by setting $V = Z^*$ and $g(V) = 1\{Z^* = z_k\}$. Unlike (6), however, it is not possible to write (F.16) as a linear inequality constraint with known slope for a general $g(V)$. On the other hand, (F.16) does satisfy

Assumption 4.2. Therefore, Theorem 4.3 implies that the "plug-in" estimator is still efficient for estimating $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ for any $\mathbb{C}$ satisfying Assumption 4.1.

EXAMPLE 2.3—Pricing Kernel:

For this example, we set $X = (Y, Z', U')'$ with $Y \in \mathbf{R}$, $Z \in \mathbf{R}^{d_Z}$, and $U \in \mathbf{R}^{d_Z}$, and hence $\mathcal{X} \subseteq \mathbf{R} \times \mathbf{R}^{d_Z} \times \mathbf{R}^{d_Z}$. Recall $\theta = (\rho, \gamma)' \in \mathbf{R}^2$, and to ensure that the identified set is bounded, we impose the constraints $0 \le \rho \le \bar{\rho}$ and $0 \le \gamma \le \bar{\gamma}$ for some $\bar{\gamma} > 0$ and $\bar{\rho} > 0$. Formally, for a parameter space Θ, the identified set is given by

$$\Theta_0(Q) \equiv \left\{\theta \in \Theta : \int \left(\frac{y^{-\gamma} z}{1+\rho} - u\right) dQ(x) \le 0 \text{ and } \theta \in [0, \bar{\rho}] \times [0, \bar{\gamma}]\right\}. \tag{F.17}$$

To map this example into (2) and (3), we let A, $m_S : \mathcal{X} \times \Theta \to \mathbf{R}^{d_Z}$, and $F_S : \mathbf{R}^{d_Z} \to \mathbf{R}^{d_Z+4}$ be given by

$$m_S(x, \theta) = \frac{y^{-\gamma} z}{1+\rho} - u, \quad F_S(v) = \left(v', -\bar{\rho}, 0, -\bar{\gamma}, 0\right)', \tag{F.18}$$

$$A' = \begin{bmatrix} 0'_{d_Z} & 1 & -1 & 0 & 0 \\ 0'_{d_Z} & 0 & 0 & 1 & -1 \end{bmatrix},$$

where 0_{d_Z} stands for $0 \in \mathbf{R}^{d_Z}$. Given this notation, the constraints $1 \le i \le d_Z$ correspond to (7), while the restriction $\theta \in [0, \bar{\rho}] \times [0, \bar{\gamma}]$ is imposed in the constraints $d_Z + 1 \le i \le d_Z + 4$. Therefore, we obtain the representation

$$\Theta_0(Q) = \left\{\theta \in \Theta : A\theta + F_S\left(\int m_S(x, \theta)\, dQ(x)\right) \le 0\right\}. \tag{F.19}$$

The following conditions are sufficient for verifying Assumptions 3.2–3.6 in Example 2.3.

ASSUMPTION F.5: (i) $\mathcal{X} \subseteq [\varepsilon_0, \infty) \times \mathbf{R}_+^{d_Z} \times \mathbf{R}^{d_Z}$ for some $\varepsilon_0 > 0$; (ii) $\mathcal{X}$ is compact; (iii) $\Theta \equiv [-1/2, 2\bar{\rho}] \times [-1/2, 2\bar{\gamma}]$.

ASSUMPTION F.6: (i) $E[\frac{Y^{-\gamma} Z}{1+\rho} - U] \le 0$ for some $\theta \in [0, \bar{\rho}] \times [0, \bar{\gamma}]$; (ii) $P(Z^{(i)} > 0) > 0$ for all $1 \le i \le d_Z$; (iii) for all $(\rho, \gamma)' = \theta \in \Theta_0(P)$, and $\{i, j\} \subseteq \mathcal{A}(\theta, P)$ with $1 \le i < j \le d_Z + 2$, $E[Y^{-\gamma}(Z^{(i)} - \pi_{i,j} Z^{(j)}) \log(Y)] \neq 0$, where $\pi_{i,j} = E[U^{(i)}]/E[U^{(j)}]$ if $j \le d_Z$ and $\pi_{i,j} = 0$ otherwise; (iv) $\#\mathcal{A}(\theta, P) \le 2$ for all $\theta \in \Theta_0(P)$.

Assumption F.5(i) requires Y, the ratio of future over current consumption, to be bounded away from zero. Together with compactness of $\mathcal{X} \times \Theta$, Assumption F.5(i) ensures $m : \mathcal{X} \times \Theta \to \mathbf{R}^{2d_Z+4}$ is bounded and differentiable, as required by Assumption 3.4. The constraint $\theta \in [0, \bar{\rho}] \times [0, \bar{\gamma}]$ can be interpreted as imposing restrictions defining the parameter space of interest (see Remark 3.4). However, our arguments require regularity of m in a neighborhood of $\Theta_0(P)$, and for this reason Assumption 3.6(i) further demands that we may define a set Θ such that $\Theta_0(P) \subset \Theta^o$. In this example, this is easily accomplished through Assumption F.5(iii); alternatively, for example, we could have set $\Theta = [-\delta, \bar{\rho} + \delta] \times [-\delta, \bar{\gamma} + \delta]$ for any $0 < \delta < 1$. Assumption 3.6(i) implies $\Theta_0(P) \neq 0$, and is satisfied if the model is properly specified. In turn, Assumption F.6(ii) is necessary for $\theta \mapsto F_S^{(i)}(\int m_S(x, \theta)\, dP(x))$ to be strictly convex for $1 \leq i \leq d_Z$. Finally, Assumption 3.6(iii)–(iv) is equivalent to Assumption 3.6(iv) in this model. Unfortunately, unlike in the linear models of Examples 2.1 and 2.2, the gradients of constraints $1 \leq i \leq d_Z$ depend on P, and as a result the requirement on P is more complex.

PROPOSITION F.2: *In Example* 2.3, *Assumptions* F.5 *and* F.6 *imply Assumptions* 3.2–3.6.

PROOF: Assumption 3.2 is implied by Assumption F.5(iii), while Assumption 3.3 has already been verified in (F.18) and (F.19). Moreover, since $y \geq \varepsilon_0 > 0$ for all $x \in \mathcal{X}$ and $\rho \geq -1/2$ for all $(\rho, \gamma)' = \theta \in \Theta$, and $\mathcal{X} \times \Theta$ is compact by Assumption F.5(i)–(ii), it also follows that $m_S(x, \theta)$ is uniformly bounded on $(x, \theta) \in \mathcal{X} \times \Theta$. Therefore, $m(x, \theta) = (m_S(x, \theta)', \theta' A')'$ implies that Assumption 3.4(i) also holds. Next, note by direct calculation that

$$\nabla_\theta m_S(x, \theta) = \left[-\frac{y^{-\gamma} z}{(1+\rho)^2} \quad -\frac{y^{-\gamma} \log(y) z}{(1+\rho)} \right], \tag{F.20}$$

and hence since $\rho \geq -1/2$ and $y \geq \varepsilon_0$ by Assumptions F.5(i) and F.5(iii), it follows that $(x, \theta) \mapsto \nabla_\theta m_S(x, \theta)$ is uniformly bounded in $\mathcal{X} \times \Theta$. Assumption 3.4(ii) then follows from $\nabla_\theta m(x, \theta) = (\nabla_\theta m_S(x, \theta)', A')'$. Moreover, (F.20) further implies $(\theta, x) \mapsto \nabla_\theta m(x, \theta)$ is continuous on $\mathcal{X} \times \Theta$. However, by compactness of $\mathcal{X} \times \Theta$, $(\theta, x) \mapsto \nabla_\theta m(x, \theta)$ is uniformly continuous, and therefore $\theta \mapsto \nabla_\theta m(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$, verifying Assumption 3.4(iii). Finally, employing $m(x, \theta) = (m_S(x, \theta)', \theta' A)'$ and $F(\int m(x, \theta)\, dQ(x)) = A\theta + F_S(\int m_S(x, \theta)\, dQ(x))$, we obtain

$$\nabla F(v) = \begin{bmatrix} I_{d_Z} & \vdots\, I_{d_Z+4} \\ 0_{4,d_Z} & \end{bmatrix}, \tag{F.21}$$

where I_k denotes the $k \times k$ identity matrix, and $0_{4,d_Z}$ is a $4 \times d_Z$ matrix of zeroes. From (F.21), it follows that Assumption 3.5(i)–(ii) holds with $V_0 = \mathbf{R}^{2d_Z+4}$.

To verify Assumption 3.6, first observe that Assumption F.2(i) directly imposes $\Theta_0(P) \neq \emptyset$. Moreover, since $\Theta_0(P) \subseteq [0, \bar{\rho}] \times [0, \bar{\gamma}] \subset (-1/2, 2\bar{\rho}) \times (-1/2, 2\bar{\gamma}) = \Theta^o$ by Assumption F.5(iii), it follows that Assumption 3.6(i) holds. To verify Assumption 3.6(ii), first note that by (F.18), $\mathcal{S}_i = \{1, 2\}$ for $1 \le i \le d_Z$ and $\mathcal{S}_i = \emptyset$ for $d_Z + 1 \le i \le d_Z + 4$. Thus, we need only show that $\theta \mapsto \int m_S^{(i)}(x, \theta)\, dQ(x)$ is strictly convex for all $1 \le i \le d_Z$ and Q in a suitable neighborhood of P. To this end, first exploit that $\rho \ge -1/2$ for all $(\rho, \gamma)' \in \Theta$ and $y \ge \varepsilon_0$ for all $x \in \mathcal{X}$ to deduce that

$$\text{(F.22)} \qquad \nabla_\theta^2 m_S^{(i)}(x, \theta) = \begin{bmatrix} \dfrac{2y^{-\gamma} z^{(i)}}{(1+\rho)^3} & \dfrac{y^{-\gamma}\log(y) z^{(i)}}{(1+\rho)^2} \\ \dfrac{y^{-\gamma}\log(y) z^{(i)}}{(1+\rho)^2} & \dfrac{y^{-\gamma}\log^2(y) z^{(i)}}{(1+\rho)} \end{bmatrix},$$

for any $(x, \theta) \in \mathcal{X} \times \Theta$ and $1 \le i \le d_Z$. By (F.22), $\nabla_\theta^2 m_S^{(i)}(x, \theta)$ is positive definite for any $x \in \mathcal{X}$ such that $z^{(i)} > 0$. Hence, since $z^{(i)} \ge 0$ on $\mathcal{X}$, and $m_S^{(i)}(x, \theta) = -u$ whenever $z^{(i)} = 0$, we conclude that, for any $\lambda \in (0, 1)$ and $1 \le i \le d_Z$,

$$\text{(F.23)} \qquad \int m_S^{(i)}\big(x, \lambda\theta_1 + (1-\lambda)\theta_2\big)\, dQ(x) < \lambda \int m_S^{(i)}(x, \theta_1)\, dQ(x) + (1-\lambda) \int m_S^{(i)}(x, \theta_2)\, dQ(x),$$

provided that $Q \in \mathbf{M}$ satisfies $Q(Z^{(i)} > 0) > 0$. However, by Assumption 3.6(ii), $P(Z^{(i)} > 0) > 0$ for all $1 \le i \le d_Z$. Hence, for each $1 \le i \le d_Z$, there exists a neighborhood $N_i(P) \subseteq \mathbf{M}$ in the τ-topology such that $Q(Z^{(i)} > 0) > 0$ for all $Q \in N_i(P)$. Therefore, by (F.23), Assumption 3.6(ii) then holds with $N(P) = \bigcap_i N_i(P)$. In turn, Assumption 3.6(iii) trivially holds since $V_0 = \mathbf{R}^{2d_Z+4}$. Finally, to verify Assumption 3.6(iv), first note

$$\text{(F.24)} \qquad \nabla F\left(\int m(x, \theta)\, dP(x)\right) \int \nabla_\theta m(x, \theta)\, dP(x) = \begin{bmatrix} -\displaystyle\int \frac{y^{-\gamma} z'}{(1+\rho)^2}\, dP(x) & 1 & -1 & 0 & 0 \\ -\displaystyle\int \frac{y^{-\gamma}\log(y) z'}{(1+\rho)}\, dP(x) & 0 & 0 & 1 & -1 \end{bmatrix}'$$

by direct calculation and (F.21). Since $P(Z^{(i)} > 0) > 0$ for all $1 \le i \le d_Z$ and $y \ge \varepsilon_0 > 0$ for all $x \in \mathcal{X}$, we must have $E[Y^{-\gamma} Z^{(i)}] > 0$. Therefore, $\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \int \nabla_\theta m(x, \theta)\, dP(x) \ne 0$ for all $1 \le i \le d_Z$, and thus $\{\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \int \nabla_\theta m(x, \theta)\, dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent if $\mathcal{A}(\theta, P)$ is either empty or singleton valued. Hence, by Assumption 3.6(iv), we need only consider the case $\mathcal{A}(\theta, P) = \{i, j\}$ with $i \ne j$. However, note that if $j \in \{d_Z + 3, d_Z + 4\}$, then $i \le d_Z + 2$ (since the $d_Z + 3$ and $d_Z + 4$ constraints cannot simultaneously bind), and, by (F.24) and $E[Y^{-\gamma} Z^{(k)}] > 0$, Assumption 3.6(iv) is satisfied. Finally, for the case $1 \leq i < j \leq d_Z + 2$, Assumption 3.6(iv) follows by direct calculation, Assumption 3.6(iii), and exploiting that if $i \in \mathcal{A}(\theta, P)$ and $i \leq d_Z$, then $E[Y^{-\gamma} Z^{(i)}] = (1+\rho)E[U^{(i)}]$. *Q.E.D.*

EXAMPLE 2.4—Participation Constraint:
In order to write this example in the form of (2) and (3), let $X = (C, W, L, Z')'$ with $(C, W, L) \in \mathbf{R}^3_+$ and $Z \in \mathbf{R}^{d_Z}_+$. We denote the parameter $\theta = (\alpha, \beta)' \in \mathbf{R}^2$ and we ensure $\Theta_0(P)$ is bounded by imposing the constraints $0 \leq \alpha \leq \bar{\alpha}$ and $0 \leq \beta \leq \bar{\beta}$ with $\bar{\alpha} > 0$ and $\bar{\beta} > 0$. For a parameter space Θ, then define the identified set

$$\text{(F.25)} \quad \Theta_0(Q) \equiv \left\{ \theta \in \Theta : \int \left(\frac{w}{c - \alpha} - \frac{\beta}{l} \right) z \, dQ(x) \leq 0 \text{ and } \theta \in [0, \bar{\alpha}] \times [0, \bar{\beta}] \right\}.$$

Further let $m_S(x, \theta) = (z'w/(c - \alpha), z'/l)'$ and define a $(d_Z + 4) \times 2$ matrix A and $F_S : \mathbf{R}^{2d_Z} \to \mathbf{R}^{d_Z+4}$ by

$$\text{(F.26)} \quad F_S^{(i)}(v) = \begin{cases} \dfrac{v^{(i)}}{v^{(d_Z+i)}}, & \text{if } 1 \leq i \leq d_Z, \\ -\bar{\alpha}, & \text{if } i = d_Z + 1, \\ 0, & \text{if } i \in \{d_Z + 2, d_Z + 4\}, \\ -\bar{\beta}, & \text{if } i = d_Z + 3, \end{cases}$$

$$A' = \begin{bmatrix} 0'_{d_Z} & 1 & -1 & 0 & 0 \\ -1'_{d_Z} & 0 & 0 & 1 & -1 \end{bmatrix},$$

where 1_{d_Z} is a vector of ones in $\mathbf{R}^{d_Z}$, and recall that 0_{d_Z} denotes $0 \in \mathbf{R}^{d_Z}$. Thus, for $1 \leq i \leq d_Z$, we obtain the constraint

$$\text{(F.27)} \quad F^{(i)}\left(\int m(x, \theta) \, dP(x) \right) = -\beta + \frac{E[WZ^{(i)}/(C - \alpha)]}{E[Z^{(i)}/L]},$$

while constraints $d_Z + 1 \leq i \leq d_Z + 4$ impose $\theta \in [0, \bar{\alpha}] \times [0, \bar{\beta}]$. Given this notation, we may then rewrite

$$\text{(F.28)} \quad \Theta_0(Q) = \left\{ \theta \in \Theta : A\theta + F_S\left(\int m_S(x, \theta) \, dQ(x) \right) \leq 0 \right\}.$$

Assumptions F.7 and F.8 impose sufficient conditions for verifying Assumptions 3.2–3.6.

ASSUMPTION F.7: (i) $\mathcal{X} \subseteq [\varepsilon_0, \infty) \times \mathbf{R}_+ \times [\varepsilon_1, +\infty) \times \mathbf{R}^{d_Z}_+$ for some $\varepsilon_0 > \bar{\alpha}$ and $\varepsilon_1 > 0$; (ii) $\mathcal{X}$ is compact; (iii) $\Theta \equiv [-\delta_0, \bar{\alpha} + \delta_0] \times [-\delta_0, \bar{\beta} + \delta_0]$ for some $0 < \delta_0 < (\varepsilon_0 - \bar{\alpha})$.

ASSUMPTION F.8: (i) $E[(\frac{W}{C-\alpha} - \frac{\beta}{L})Z] \leq 0$ for some $\theta \in [0, \bar{\alpha}] \times [0, \bar{\beta}]$; (ii) $P(WZ^{(i)} > 0) > 0$ for all $1 \leq i \leq d_Z$; (iii) for all $(\alpha, \beta)' = \theta \in \Theta_0(P)$, and $\{i, j\} \subseteq \mathcal{A}(\theta, P)$ with $1 \leq i < j \leq d_Z$, $E[\frac{W}{(C-\alpha)^2}(Z^{(i)} - \pi_{i,j}Z^{(j)})] \neq 0$, where $\pi_{i,j} = E[\frac{Z^{(i)}}{L}]/E[\frac{Z^{(j)}}{L}]$; (iv) $\#\mathcal{A}(\theta, P) \leq 2$ for all $\theta \in \Theta_0(P)$.

In Assumptions F.7(i) and F.7(iii), we impose that $C - \alpha$ and L be bounded away from zero, as required for $m(x, \theta)$ to be bounded and utility to remain finite (recall $u(C, L) = \log(C - \alpha) + \beta \log(L)$). As in Example 2.3, in Assumption F.7(iii) we define Θ to be an expansion of the parameter constraints $\theta \in [0, \bar{\alpha}] \times [0, \bar{\beta}]$. Assumption F.8(i) ensures $\Theta_0(P) \neq \emptyset$, while Assumption F.8(ii) is required so that constraints $1 \leq i \leq d_Z$ are strictly convex in α. Finally, Assumptions F.8(iii) and F.8(iv) are necessary and sufficient for P to satisfy Assumption 3.6(iv) in this model. As in Example 2.3, the gradients of constraints $1 \leq i \leq d_Z$ depend on P, which leads to a more complex requirement than was necessary in Examples 2.1 and 2.2.

PROPOSITION F.3: *In Example* 2.4, *Assumptions* F.7 *and* F.8 *imply Assumptions* 3.2–3.6.

PROOF: Assumption 3.2 is implied by Assumption F.7(iii), while Assumption 3.3 was already verified in (F.28). Moreover, compactness of $\mathcal{X} \times \Theta$ implies wz and z are uniformly bounded, while $c \geq \varepsilon_0 > \bar{\alpha} + \delta_0 \geq \alpha$ and $l \geq \varepsilon_1$ imply $(c - \alpha)^{-1}$ and l^{-1} are uniformly bounded as well. Therefore, $m_S(x, \theta) = (z'w/(c - \alpha), z'/l)'$ is uniformly bounded in $(x, \theta) \in \mathcal{X} \times \Theta$ and hence so is $m(x, \theta) = (m_S(x, \theta)', \theta' A')'$, which verifies Assumption 3.4(i). Similarly,

$$\nabla_\theta m_S(x, \theta) = \begin{bmatrix} \dfrac{wz}{(c - \alpha)^2} & 0_{d_z} \\ 0_{d_Z} & 0_{d_Z} \end{bmatrix} \tag{F.29}$$

is also bounded, which, together with $\nabla_\theta m(x, \theta) = (\nabla_\theta m_S(x, \theta)', A')'$, implies Assumption 3.4(ii) holds as well. In turn, by compactness of $\mathcal{X} \times \Theta$ and (F.29), $(x, \theta) \mapsto \nabla_\theta m(x, \theta)$ is uniformly continuous on $\mathcal{X} \times \Theta$ and therefore $\theta \mapsto \nabla_\theta m(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$ as demanded by Assumption 3.4(iii). Next, let $\eta_0 < \inf_k E[Z^{(k)}/L]$ and note that we may set $\eta_0 > 0$ due to Assumption F.8(ii) and $P(W \geq 0) = 1$ by definition of $\mathcal{X}$. Similarly, let $\sup_{\mathcal{X} \times \Theta} \|m(x, \theta)\| < M_0$, and note that since Assumption 3.4(ii) holds, we may set $M_0 < \infty$. Then defining

$$V_0 \equiv (-M_0, M_0)^{dz} \times (\eta_0, M_0)^{dz} \times (-M_0, M_0)^{dz+4}, \tag{F.30}$$

and noting that $F(v)$ is differentiable unless $v^{(i)} = 0$ for some $d_Z + 1 \leq i \leq 2d_Z$, it follows that Assumption 3.5(i) holds. In addition, since ∇F is continuous on the closure of V_0 and such closure is compact, it follows that Assumption 3.5(ii) holds as well.

To verify that P satisfies Assumption 3.6, first note that by Assumptions F.7(iii) and F.8(i), $\emptyset \neq \Theta_0(P) \subseteq [0, \bar{\alpha}] \times [0, \bar{\beta}] \subset (-\delta_0, \bar{\alpha}+\delta_0) \times (-\delta_0, \bar{\beta}+\delta_0) = \Theta^o$, which verifies Assumption 3.6(i). Next observe $\mathcal{S}_i = \emptyset$ for $d_Z + 1 \leq i \leq d_Z + 4$, and $\mathcal{S}_i = \{1\}$ for $1 \leq i \leq d_Z$. Therefore, to show that Assumption 3.6(ii) holds, it suffices to establish that

$$F_S^{(i)}\left(\int m(x, \theta)\, dQ(x)\right) = \frac{\displaystyle\int (wz^{(i)}/(c-\alpha))\, dQ(x)}{\displaystyle\int (z^{(i)}/l)\, dQ(x)} \tag{F.31}$$

is strictly convex in α for all Q in an appropriate neighborhood of P. However, by Assumptions F.7(i) and F.8(ii), $E[Z^{(i)}/L] > 0$ and $E[WZ^{(i)}] > 0$, and therefore there exists a neighborhood $N_i(P) \subseteq \mathbf{M}$ such that $\int (z^{(i)}/l)\, dQ(x) > 0$ and $\int wz^{(i)}\, dQ(x) > 0$ for all $Q \in N_i(P)$. Letting $N(P) = \bigcap_i N_i(P)$ and noting that $Q(C - \alpha > 0) = 1$ for all $\alpha \in [0, \bar{\alpha}]$ and $Q \in \mathbf{M}$ by Assumption F.7(i), we obtain that $\alpha \mapsto F_S^{(i)}(\int m(x, \theta)\, dQ(x))$ is indeed strictly convex for all $Q \in N(P)$, thus verifying Assumption 3.6(ii). In turn, Assumption 3.6(iii) is also satisfied by construction of V_0 in (F.30) and definitions of η_0 and M_0. Finally, note that by (F.29) and direct calculation,

$$\nabla F\left(\int m(x, \theta)\, dP(x)\right) \int \nabla_\theta m(x, \theta)\, dP(x) = \begin{bmatrix} \dfrac{E[WZ'/(C-\alpha)^2]}{E[Z'/L]} & 1 & -1 & 0 & 0 \\ -1'_{d_Z} & 0 & 0 & 1 & -1 \end{bmatrix}'. \tag{F.32}$$

Hence, since $E[WZ^{(i)}/(C-\alpha)] > 0$ and $E[Z^{(i)}/L] > 0$ for all $\alpha \in [0, \bar{\alpha}]$ and $1 \leq i \leq d_Z$ by Assumptions F.7(i) and F.8(ii), (F.32) implies $\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \int \frac{\partial}{\partial \theta^{(j)}} m(x, \theta)\, dP(x) \neq 0$ for any $1 \leq i \leq d_Z$ and $j \in \{1, 2\}$. As a result, it follows that $\{\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \int \nabla_\theta m(x, \theta)\, dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent whenever $\mathcal{A}(\theta, P)$ is empty or a singleton, and also when $\{i, j\} = \mathcal{A}(\theta, P)$ with $i < j$ and $j \geq d_Z + 1$. Therefore, by Assumption F.8(iv), to verify Assumption 3.6(iv) it only remains to consider the case $\{i, j\} = \mathcal{A}(\theta, P)$ with $j \leq d_Z$. However, in this instance, $\{\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \int \nabla_\theta m(x, \theta)\, dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent by result (F.32), Assumption F.8(iv), and direct calculation, and hence P satisfies Assumption 3.6(iv) as well. *Q.E.D.*

APPENDIX G: SIMULATION EVIDENCE

In this appendix, we assess the finite sample performance of the efficient estimator and illustrate its ease of implementation with a Monte Carlo experiment based on Example 2.1. For comparison purposes, we also include the results of employing the uniformly valid procedures proposed in Andrews and Soares (2010) and Bugni (2010).

For our design, we let $Z_i \equiv (Z_i^{(1)}, Z_i^{(2)})'$, where $Z_i^{(1)} = 1$ is a constant and $Z_i^{(2)}$ is uniformly distributed on a set $\mathcal{Z}_2$ of K equally spaced points on $[-5, 5]$. For a true parameter $\theta_0 = (1, 2)'$, we then generate Y_i according to

$$Y_i = Z_i'\theta_0 + \varepsilon_i, \quad i = 1, \ldots, n, \tag{G.1}$$

where ε_i is a standard normal random variable independent of Z_i. We assume Y_i is unobservable, but create observable upper and lower bounds $(Y_{L,i}, Y_{U,i})$ such that $Y_{L,i} \leq Y_i \leq Y_{U,i}$ almost surely. Specifically, we let

$$\begin{aligned} Y_{L,i} &= Y_i - C - V_i\big(Z_i^{(2)}\big)^2, \quad i = 1, \ldots, n, \\ Y_{U,i} &= Y_i + C + V_i\big(Z_i^{(2)}\big)^2, \quad i = 1, \ldots, n, \end{aligned} \tag{G.2}$$

where $C > 0$ and V_i is uniformly distributed on $[0, 0.2]$ independently of $(Y_i, Z_i')'$. As discussed in Example 2.1, $\Theta_0(P)$ consists of all $\theta \in \Theta$ such that $E[Y_{L,i}|Z_i] \leq Z_i'\theta \leq E[Y_{U,i}|Z_i]$ almost surely (see also (F.2)). All our reported simulation results are based on 5000 replications.

Our Monte Carlo experiment is designed to examine the robustness of the estimator to the two free parameters K and C. Since $d_F = 2K$, the constant K determines the number of constraints, while C controls the diameter of the identified set, with point identification occurring at $C = 0$—see Figure 1. Throughout our simulation study, we will examine specifications with $C \in \{0.1, 0.5, 1\}$ and $K \in \{5, 9, 15\}$, with the latter corresponding to 10, 18, and 30 moment inequalities, respectively. Heuristically, high values of K or low values of C yield specifications where P is closer to violating Assumption 3.6(iv). In such instances, we therefore expect our asymptotic results to provide a less

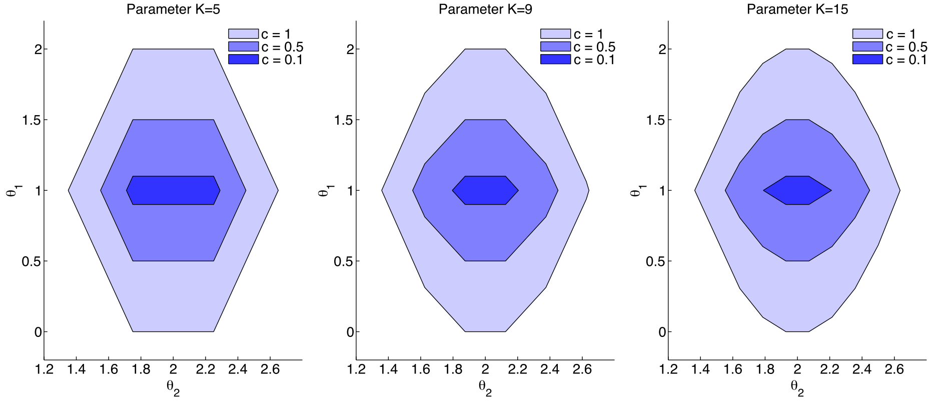

FIGURE 1.—Identified set as a function of C and K.

reliable approximation to finite sample distributions, while uniform procedures should remain accurate.

We first compare the performance of the efficient set estimator $\hat{\Theta}_n = \text{co}\{\Theta_0(\hat{P}_n)\}$ (see (20)) with that of

$$\hat{\Theta}_n(\tau_n) \equiv \left\{\theta \in \Theta : F^{(i)}\left(\int m(x,\theta)\, d\hat{P}_n(x)\right) \leq \frac{\tau_n}{\sqrt{n}} \hat{\sigma}_n^{(i)} \text{ for } i = 1, \ldots, d_F\right\}, \tag{G.3}$$

where $(\hat{\sigma}_n^{(i)})^2$ is a consistent estimator for the asymptotic variance of constraint number i.[15] Chernozhukov, Hong, and Tamer (2007) and Bugni (2010) showed that $\hat{\Theta}_n(\tau_n)$ is a consistent estimator for $\Theta_0(P)$ under the Hausdorff metric provided that $\tau_n/\sqrt{n} \downarrow 0$. Notice, in particular, that the efficient estimator $\hat{\Theta}_n$ corresponds to setting $\tau_n = 0$, and is therefore by construction always smaller than $\hat{\Theta}_n(\tau_n)$ whenever $\tau_n > 0$. This is not necessarily a favorable property, however, since an estimator that is too small may perform poorly in terms of Hausdorff distance to $\Theta_0(P)$. For example, in certain specifications, we find in many replications that $\hat{\Theta}_n(\tau_n) = \emptyset$ for values of $\tau_n \in \{0, \log(\log(n))\}$, in which case the Hausdorff distance to $\Theta_0(P)$ is set to equal infinity. Table I reports the proportion of replications for which this event occurs in each specification. As expected, the most problematic specifications are those with many moment inequalities ($K = 15$) and $\Theta_0(P)$ near point identification ($C = 0.1$).

TABLE I

PROPORTION OF SIMULATED SAMPLES WITH EMPTY SET ESTIMATORS

	$K=5$			$K=9$			$K=15$		
Sample Size	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				Estimator $\Theta_0(\hat{P}_n)$					
$n=200$	–	–	–	0.201	–	–	0.792	0.010	–
$n=500$	–	–	–	0.035	–	–	0.420	–	–
$n=1000$	–	–	–	0.003	–	–	0.152	–	–
				Estimator $\hat{\Theta}_n(\tau_n)$ With $\tau_n = \log(\log(n))$					
$n=200$	–	–	–	–	–	–	0.007	–	–
$n=500$	–	–	–	–	–	–	–	–	–
$n=1000$	–	–	–	–	–	–	–	–	–

[15]In particular, for $\bar{m}_n(\theta) \equiv \int m(x,\theta)\, d\hat{P}_n(x)$, and $\hat{\Omega}_n(\theta) \equiv \int (m(x,\theta) - \bar{m}_n(\theta))(m(x,\theta) - \bar{m}_n(\theta))'\, d\hat{P}_n(x)$, we let $(\hat{\sigma}_n^{(i)})^2 \equiv \nabla F^{(i)}(\int m(x,\theta)\, d\hat{P}_n(x))\hat{\Omega}_n(\theta)\nabla F^{(i)}(\int m(x,\theta)\, d\hat{P}_n(x))'$. It is easy to verify that $\hat{\sigma}_n^{(i)}$ does not depend on θ.

TABLE II

MEDIAN HAUSDORFF DISTANCE

	$K=5$			$K=9$			$K=15$		
Estimator	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Efficient	0.131	0.132	0.132	0.232	0.208	0.209	Inf	0.332	0.332
$\tau_n=\log(\log(n))$	0.372	0.372	0.372	0.423	0.423	0.423	0.393	0.392	0.392
$\tau_n=\log(n)$	0.941	0.941	0.941	1.138	1.138	1.138	1.226	1.226	1.226
$\tau_n=n^{1/8}$	0.414	0.414	0.414	0.476	0.476	0.476	0.455	0.455	0.455
$\tau_n=n^{1/4}$	0.702	0.702	0.702	0.838	0.838	0.838	0.879	0.879	0.879
				$n=500$					
Efficient	0.080	0.081	0.081	0.136	0.130	0.131	0.290	0.205	0.204
$\tau_n=\log(\log(n))$	0.251	0.251	0.251	0.316	0.316	0.316	0.315	0.315	0.315
$\tau_n=\log(n)$	0.692	0.692	0.692	0.890	0.890	0.890	1.021	1.021	1.021
$\tau_n=n^{1/8}$	0.285	0.285	0.285	0.362	0.362	0.362	0.371	0.371	0.371
$\tau_n=n^{1/4}$	0.542	0.542	0.542	0.696	0.696	0.696	0.783	0.783	0.783
				$n=1000$					
Efficient	0.058	0.058	0.058	0.093	0.092	0.093	0.172	0.144	0.144
$\tau_n=\log(\log(n))$	0.185	0.185	0.185	0.244	0.244	0.244	0.257	0.257	0.257
$\tau_n=\log(n)$	0.537	0.537	0.537	0.713	0.713	0.713	0.841	0.841	0.841
$\tau_n=n^{1/8}$	0.216	0.216	0.216	0.285	0.285	0.285	0.308	0.308	0.308
$\tau_n=n^{1/4}$	0.447	0.447	0.447	0.592	0.592	0.592	0.690	0.690	0.690

Table II reports the median of the Hausdorff distance between the different set estimators and $\Theta_0(P)$ across replications—see Remark G.1 for computational details. We report median, rather than mean, Hausdorff distance because $d_H(\hat{\Theta}_n(\tau_n), \Theta_0(P))$ is infinite in replications for which $\hat{\Theta}_n(\tau_n) = \emptyset$. As expected, the median Hausdorff distance decreases with sample size across all specifications and choices of τ_n. Interestingly, for $\tau_n \in \{\log(\log(n)), \log(n), n^{1/8}, n^{1/4}\}$, the performance of $\hat{\Theta}_n(\tau_n)$ is completely insensitive to the choice of C across all specifications, while the performance of the efficient estimator is only sensitive to the value of C when many moment inequalities are present ($K = 15$). In contrast, the median Hausdorff distance of all estimators deteriorates as the number of moment inequalities increases. Remarkably, across almost all specifications, the median Hausdorff distance is monotonically increasing in τ_n, with the efficient estimator outperforming all the alternative estimators.[16] The notable exception is the specification $K = 15$, $C = 0.1$, and $n = 200$, in which the median Hausdorff distance of the efficient estimator is infinite due to $\Theta_0(\hat{P}_n)$ being empty in over half the replications (see Table I).

[16]Note that for all the values of n we consider, $\log(\log(n)) < n^{1/8} < \log(n) < n^{1/4}$.

Next, we examine the performance of inferential procedures based on the semiparametric efficient estimator and compare it to that of alternative methods that are asymptotically valid uniformly in P. To this end, we first consider the construction of confidence regions $\mathcal{C}_n$ for the identified set $\Theta_0(P)$ satisfying the coverage requirement

$$\liminf_{n\to\infty} P\big(\Theta_0(P)\subseteq \mathcal{C}_n\big)\geq 1-\alpha. \tag{G.4}$$

Following the discussion in Example 5.1, we employ the efficient estimator to obtain a confidence region satisfying (G.4) by using a construction proposed in Beresteanu and Molinari (2008)—see Remark G.2 for computational details. Additionally, we also obtain confidence regions satisfying (G.4) by utilizing a criterion function based approach, as developed in Chernozhukov, Hong, and Tamer (2007) and Bugni (2010). Specifically, defining the criterion function

$$Q_n(\theta)\equiv \max_{1\leq i\leq d_F}\frac{1}{\hat{\sigma}_n^{(i)}}\left(F^{(i)}\left(\int m(x,\theta)\,d\hat{P}_n(x)\right)\right)_+, \tag{G.5}$$

we examine confidence regions of the form $\mathrm{CS}_n(\tau_n)\equiv\{\theta\in\Theta: Q_n(\theta)\leq \hat{c}^B_{1-\alpha}(\tau_n)/\sqrt{n}\}$, where $\hat{c}^B_{1-\alpha}(\tau_n)$ is the critical value proposed in Bugni (2010)—see Remark G.3. Employing the maximum, rather than the sum, across constraints in defining Q_n implies $\mathrm{CS}_n(\tau_n)$ is a convex polygon, which greatly simplifies our computations. All bootstrap procedures employed 200 replications in computing critical values.

Table III reports the coverage probabilities of the different confidence regions under alternative values of (n, K, C) for a nominal coverage of 0.95. A confidence region based on the efficient estimator is considered to have failed to cover $\Theta_0(P)$ in any replication for which $\Theta_0(\hat{P}_n)=\emptyset$. Similarly, the criterion based confidence region is considered to have failed to cover $\Theta_0(P)$ whenever $\hat{\Theta}_n(\tau_n)=\emptyset$—see Remark G.3. As in Table II, the performance of the confidence region based on the efficient estimator is more sensitive to K than to C. In specifications with 10 moment inequalities ($K=5$), the actual coverage is always close to its nominal level, while under 30 moment inequalities ($K=15$), size distortions upwards of 5% remain even for $n=1000$. Unsurprisingly, the most severe undercoverage occurs in specifications for which $\Theta_0(\hat{P}_n)=\emptyset$ in a large number of replications ($K=15, C=0.1$). In contrast, the criterion based confidence regions have actual coverage above the nominal level for all specifications. The coverage probability is closest to the nominal level under 10 moment inequalities ($K=5$), but can be quite conservative for larger values of the slackness parameter τ_n ($\tau_n\in\{\log(n), n^{1/4}\}$).

In Table IV, we report the median Hausdorff distance between the different confidence regions and the identified set $\Theta_0(P)$. For specifications in

TABLE III

SET CONFIDENCE REGION COVERAGE PROBABILITY—NOMINAL COVERAGE = 0.95

Procedure	$K=5$			$K=9$			$K=15$		
	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Efficient	0.945	0.942	0.940	0.790	0.913	0.895	0.208	0.885	0.820
B. $\tau_n = \log(\log(n))$	0.980	0.984	0.986	0.990	0.992	0.992	0.989	0.997	0.998
B. $\tau_n = \log(n)$	0.998	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
B. $\tau_n = n^{1/8}$	0.983	0.986	0.987	0.993	0.993	0.994	0.994	0.998	0.999
B. $\tau_n = n^{1/4}$	0.994	0.995	0.995	0.999	0.999	0.999	1.000	1.000	1.000
				$n=500$					
Efficient	0.950	0.940	0.940	0.946	0.916	0.916	0.573	0.879	0.870
B. $\tau_n = \log(\log(n))$	0.967	0.972	0.978	0.983	0.982	0.982	0.986	0.988	0.990
B. $\tau_n = \log(n)$	0.994	0.993	0.995	0.999	0.999	0.998	1.000	1.000	1.000
B. $\tau_n = n^{1/8}$	0.971	0.975	0.979	0.987	0.984	0.986	0.990	0.991	0.991
B. $\tau_n = n^{1/4}$	0.989	0.989	0.990	0.998	0.998	0.997	0.999	1.000	1.000
				$n=1000$					
Efficient	0.959	0.946	0.946	0.975	0.926	0.925	0.829	0.891	0.889
B. $\tau_n = \log(\log(n))$	0.969	0.971	0.979	0.983	0.981	0.980	0.981	0.983	0.982
B. $\tau_n = \log(n)$	0.991	0.993	0.994	0.998	0.998	0.997	1.000	1.000	0.999
B. $\tau_n = n^{1/8}$	0.970	0.973	0.981	0.987	0.984	0.983	0.985	0.986	0.986
B. $\tau_n = n^{1/4}$	0.989	0.989	0.992	0.997	0.996	0.994	0.999	0.999	0.998

which all confidence regions control size, the median Hausdorff distance of the confidence region based on the efficient estimator is always smaller than that of its competitors. These results suggest that while the criterion based confidence regions can deliver uniform size control, they can also underperform when our asymptotic results provide an accurate approximation to finite sample distributions. Finally, in Table V, we tabulate the median computation time in seconds for each confidence region. The computational time of all approaches is small, but longest for the confidence region based on the efficient estimator. It is worth noting that the Lagrange multipliers $\lambda(p, \hat{P}_n)$ and maximizers $\hat{\theta}(p)$ needed to construct $G_n^*(p)$ (as in (23)) are by-products of computing $\nu(p, \Theta_0(\hat{P}_n))$. As a result, simulating the distribution of G_n^* only requires sampling $\{W_i\}_{i=1}^n$, which significantly reduces computation time relative to a procedure that recomputes the support function in each bootstrap iteration.

We further evaluate the size and power of the test based on $J_n(\theta)$ (see (28)) for the null hypothesis:

$$H_0 : \theta \in \Theta_0(P), \quad H_1 : \theta \notin \Theta_0(P). \tag{G.6}$$

TABLE IV

SET CONFIDENCE REGION MEDIAN HAUSDORFF DISTANCE—NOMINAL COVERAGE 0.95

	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Efficient	0.439	0.430	0.430	0.639	0.502	0.498	Inf	0.542	0.524
B. $\tau_n=\log(\log(n))$	0.566	0.576	0.585	0.826	0.855	0.871	1.504	1.645	1.724
B. $\tau_n=\log(n)$	0.710	0.718	0.727	1.245	1.259	1.275	3.257	3.359	3.413
B. $\tau_n=n^{1/8}$	0.577	0.585	0.594	0.855	0.882	0.897	1.629	1.764	1.834
B. $\tau_n=n^{1/4}$	0.645	0.651	0.660	1.058	1.075	1.092	2.505	2.593	2.672
				$n=500$					
Efficient	0.276	0.273	0.273	0.398	0.349	0.349	0.740	0.371	0.373
B. $\tau_n=\log(\log(n))$	0.328	0.334	0.344	0.468	0.481	0.488	0.580	0.600	0.609
B. $\tau_n=\log(n)$	0.384	0.387	0.392	0.597	0.603	0.608	0.862	0.876	0.883
B. $\tau_n=n^{1/8}$	0.332	0.339	0.347	0.478	0.490	0.496	0.600	0.618	0.626
B. $\tau_n=n^{1/4}$	0.366	0.369	0.374	0.551	0.557	0.562	0.759	0.773	0.779
				$n=1000$					
Efficient	0.195	0.193	0.193	0.282	0.262	0.262	0.388	0.282	0.283
B. $\tau_n=\log(\log(n))$	0.226	0.227	0.237	0.326	0.335	0.341	0.394	0.405	0.411
B. $\tau_n=\log(n)$	0.257	0.258	0.261	0.389	0.392	0.394	0.514	0.519	0.523
B. $\tau_n=n^{1/8}$	0.228	0.230	0.239	0.333	0.341	0.345	0.404	0.414	0.419
B. $\tau_n=n^{1/4}$	0.250	0.252	0.254	0.372	0.376	0.378	0.480	0.485	0.489

In order to make size control nontrivial, we let θ be a boundary point of $\Theta_0(P)$. In particular, for the vectors

$$\theta_F \equiv \big(\nu\big((1,0),\Theta_0(P)\big),0\big)', \quad \theta_K \equiv \big(0,\nu\big((0,1),\Theta_0(P)\big)\big)', \tag{G.7}$$

we consider the hypothesis testing problem in (G.6) when $\theta\in\{\theta_F,\theta_K\}$. Notice that θ_F and θ_K are respectively points in a "flat face" and at a "kink" of $\Theta_0(P)$ for all values of (C,K) (see Figure 1). Thus, θ_F is supported by a unique hyperplane while θ_K is supported by multiple hyperplanes, which implies that Theorem 5.3 applies to the former but not the latter. For comparison purposes, we also examine the performance of the generalized moment selection procedure developed in Andrews and Soares (2010). Specifically, for $\theta\in\{\theta_F,\theta_K\}$, we consider a test that rejects the null hypothesis in (G.6) whenever $\sqrt{n}Q_n(\theta) > \hat{c}^{AS}_{1-\alpha}(\theta)$ for a bootstrap critical value $\hat{c}^{AS}_{1-\alpha}(\theta)$—see Remark G.4. Both procedures require a choice of slackness parameter (see (30)), which we select from the set $\{\log(\log(n)),\log(n),n^{1/8},n^{1/4}\}$.

Tables VI and VII report the actual size of tests of (G.6) for a nominal size of 0.05 and $\theta\in\{\theta_F,\theta_K\}$. For tests based on the efficient estimator, we considered the null hypothesis in (G.6) to be rejected in any replication for which $\Theta_0(\hat{P}_n)=\emptyset$. The performance of the tests for (G.6) when $\theta=\theta_F$ are similar

TABLE V

MEDIAN CONFIDENCE REGION COMPUTATION TIME IN SECONDS

	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Efficient	2.528	2.683	2.751	3.741	4.061	4.319	5.824	6.138	6.514
B. $\tau_n=\log(\log(n))$	1.997	1.946	1.917	2.308	2.436	2.535	3.023	3.342	3.501
B. $\tau_n=\log(n)$	1.925	1.890	1.907	2.434	2.492	2.545	3.311	3.431	3.495
B. $\tau_n=n^{1/8}$	1.995	1.944	1.917	2.335	2.455	2.546	3.086	3.368	3.519
B. $\tau_n=n^{1/4}$	1.976	1.923	1.921	2.421	2.503	2.572	3.281	3.452	3.542
				$n=500$					
Efficient	2.577	2.691	2.730	3.805	4.191	4.493	5.867	6.397	6.839
B. $\tau_n=\log(\log(n))$	2.086	2.002	1.936	2.420	2.554	2.660	3.277	3.538	3.741
B. $\tau_n=\log(n)$	2.007	1.947	1.919	2.543	2.607	2.678	3.536	3.627	3.749
B. $\tau_n=n^{1/8}$	2.082	1.998	1.936	2.442	2.565	2.670	3.344	3.565	3.758
B. $\tau_n=n^{1/4}$	2.049	1.983	1.933	2.534	2.617	2.702	3.528	3.644	3.797
				$n=1000$					
Efficient	2.587	2.649	2.654	3.758	4.203	4.484	5.742	6.383	6.952
B. $\tau_n=\log(\log(n))$	2.174	2.055	1.979	2.532	2.656	2.754	3.366	3.626	3.767
B. $\tau_n=\log(n)$	2.086	2.010	1.947	2.641	2.709	2.786	3.613	3.729	3.802
B. $\tau_n=n^{1/8}$	2.166	2.049	1.980	2.551	2.673	2.769	3.427	3.668	3.804
B. $\tau_n=n^{1/4}$	2.124	2.034	1.963	2.653	2.728	2.812	3.613	3.760	3.854

to those of the confidence regions for $\Theta_0(P)$ (Table III). In particular, the test based on the efficient estimator provides accurate size control under ten moment inequalities ($K=5$), but can fail to do so under 30 moment inequalities ($K=15$). With the exception of those specifications in which $\Theta_0(\hat{P}_n)=\emptyset$ in a significant number of replications, however, the size distortions are not as severe as those in Table III. In contrast, the test of Andrews and Soares (2010) always provides adequate size control, though it can sometimes be severely conservative, for instance for $K=15$ and $C=0.1$. The patterns when $\theta=\theta_K$ are similar, though all tests have a weakly lower rejection rate than when $\theta=\theta_F$ in a majority of the specifications. As a result, for larger values of κ_n ($\kappa_n\in\{\log(n), n^{1/4}\}$), the test based on the efficient estimator delivers adequate size control in all specifications except those for which $\Theta_0(\hat{P}_n)=\emptyset$ in a large proportion of replications (see Table I).

In order to evaluate the local power of the proposed tests, we further test (G.6) when θ is of the form

$$\theta=\theta_C+\frac{h}{\sqrt{n}}\theta_A, \tag{G.8}$$

TABLE VI

EMPIRICAL SIZE $H_0: \theta_F \in \Theta_0(P)$ (ON FLAT FACE)—NOMINAL SIZE = 0.05

	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Eff. $\kappa_n=\log(\log(n))$	0.037	0.055	0.056	0.205	0.066	0.073	0.792	0.113	0.146
Eff. $\kappa_n=\log(n)$	0.034	0.054	0.056	0.204	0.057	0.067	0.792	0.089	0.134
Eff. $\kappa_n=n^{1/8}$	0.036	0.054	0.056	0.205	0.065	0.072	0.792	0.110	0.144
Eff. $\kappa_n=n^{1/4}$	0.035	0.054	0.056	0.204	0.058	0.068	0.792	0.094	0.136
A.S. $\tau_n=\log(\log(n))$	0.040	0.040	0.039	0.012	0.016	0.015	0.004	0.006	0.007
A.S. $\tau_n=\log(n)$	0.011	0.017	0.019	0.006	0.008	0.009	0.003	0.004	0.004
A.S. $\tau_n=n^{1/8}$	0.039	0.039	0.039	0.012	0.014	0.014	0.003	0.006	0.007
A.S. $\tau_n=n^{1/4}$	0.018	0.024	0.026	0.007	0.011	0.011	0.003	0.004	0.005
				$n=500$					
Eff. $\kappa_n=\log(\log(n))$	0.040	0.052	0.052	0.045	0.053	0.053	0.421	0.090	0.093
Eff. $\kappa_n=\log(n)$	0.034	0.052	0.052	0.040	0.047	0.047	0.420	0.076	0.082
Eff. $\kappa_n=n^{1/8}$	0.039	0.052	0.052	0.044	0.052	0.052	0.420	0.089	0.092
Eff. $\kappa_n=n^{1/4}$	0.035	0.052	0.052	0.040	0.048	0.048	0.420	0.079	0.084
A.S. $\tau_n=\log(\log(n))$	0.049	0.050	0.049	0.017	0.022	0.022	0.012	0.018	0.017
A.S. $\tau_n=\log(n)$	0.016	0.027	0.027	0.007	0.012	0.011	0.006	0.008	0.010
A.S. $\tau_n=n^{1/8}$	0.049	0.050	0.049	0.016	0.021	0.021	0.011	0.016	0.016
A.S. $\tau_n=n^{1/4}$	0.024	0.043	0.041	0.008	0.014	0.013	0.008	0.011	0.012
				$n=1000$					
Eff. $\kappa_n=\log(\log(n))$	0.048	0.049	0.049	0.054	0.056	0.056	0.189	0.105	0.105
Eff. $\kappa_n=\log(n)$	0.038	0.049	0.049	0.020	0.054	0.054	0.154	0.082	0.083
Eff. $\kappa_n=n^{1/8}$	0.030	0.048	0.048	0.011	0.048	0.048	0.152	0.070	0.071
Eff. $\kappa_n=n^{1/4}$	0.037	0.048	0.048	0.017	0.054	0.054	0.154	0.080	0.081
A.S. $\tau_n=\log(\log(n))$	0.050	0.050	0.048	0.026	0.027	0.028	0.016	0.020	0.051
A.S. $\tau_n=\log(n)$	0.023	0.045	0.043	0.008	0.011	0.011	0.006	0.009	0.020
A.S. $\tau_n=n^{1/8}$	0.050	0.050	0.048	0.020	0.023	0.024	0.014	0.018	0.009
A.S. $\tau_n=n^{1/4}$	0.024	0.049	0.048	0.008	0.014	0.015	0.007	0.011	0.017

where $\theta_C \in \{\theta_F, \theta_K\}$, and $\theta_A = (1,0)'$ if $\theta_C = \theta_F$ and $\theta_C = (0,1)'$ otherwise. It can be verified by direct calculation that $h/\sqrt{n} = \inf_{\tilde{\theta}\in\Theta_0(P)} \|\theta - \tilde{\theta}\|$ whenever $h \geq 0$, and hence h controls the distance of the local alternative to the identified set. Tables VIII and IX report rejection probabilities for tests with a nominal size of 0.05. We focus on specifications with $K \in \{5, 9\}$ so that both tests provide adequate size control, and ignore specifications with $n = 500$ for conciseness. Notice that results with $h = 0$ correspond to the actual size of the test. For local deviations away from $\theta = \theta_F$ (Table VIII), the test based on the efficient estimator is more powerful than its competitors in almost all specifications, and the pattern is robust to the choice of slackness parameters. Interestingly, all tests are more powerful in detecting local deviations away from $\theta = \theta_K$ (Table IX) than from $\theta = \theta_F$. However, in this instance, the tests are

TABLE VII

EMPIRICAL SIZE $H_0: \theta_K \in \Theta_0(P)$ (ON KINK)—NOMINAL SIZE = 0.05

Procedure	$K=5$			$K=9$			$K=15$		
	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
				$n=200$					
Eff. $\kappa_n=\log(\log(n))$	0.028	0.044	0.062	0.214	0.056	0.090	0.796	0.110	0.149
Eff. $\kappa_n=\log(n)$	0.024	0.005	0.024	0.206	0.017	0.017	0.793	0.046	0.046
Eff. $\kappa_n=n^{1/8}$	0.028	0.039	0.057	0.213	0.050	0.080	0.795	0.098	0.137
Eff. $\kappa_n=n^{1/4}$	0.025	0.013	0.037	0.208	0.025	0.035	0.793	0.060	0.073
A.S. $\tau_n=\log(\log(n))$	0.028	0.035	0.027	0.016	0.013	0.017	0.003	0.003	0.005
A.S. $\tau_n=\log(n)$	0.020	0.019	0.023	0.010	0.010	0.013	0.002	0.002	0.004
A.S. $\tau_n=n^{1/8}$	0.027	0.032	0.026	0.015	0.012	0.015	0.003	0.002	0.005
A.S. $\tau_n=n^{1/4}$	0.023	0.020	0.025	0.012	0.010	0.013	0.003	0.002	0.004
				$n=500$					
Eff. $\kappa_n=\log(\log(n))$	0.010	0.048	0.055	0.060	0.047	0.087	0.435	0.089	0.119
Eff. $\kappa_n=\log(n)$	0.009	0.007	0.029	0.045	0.012	0.017	0.426	0.024	0.029
Eff. $\kappa_n=n^{1/8}$	0.010	0.043	0.052	0.058	0.040	0.078	0.434	0.079	0.106
Eff. $\kappa_n=n^{1/4}$	0.009	0.016	0.037	0.048	0.015	0.033	0.428	0.034	0.047
A.S. $\tau_n=\log(\log(n))$	0.026	0.045	0.029	0.023	0.020	0.028	0.018	0.016	0.017
A.S. $\tau_n=\log(n)$	0.020	0.017	0.024	0.013	0.010	0.019	0.011	0.011	0.013
A.S. $\tau_n=n^{1/8}$	0.026	0.044	0.028	0.022	0.019	0.026	0.017	0.015	0.017
A.S. $\tau_n=n^{1/4}$	0.023	0.023	0.024	0.016	0.011	0.020	0.012	0.012	0.015
				$n=1000$					
Eff. $\kappa_n=\log(\log(n))$	0.006	0.050	0.054	0.037	0.047	0.082	0.197	0.089	0.107
Eff. $\kappa_n=\log(n)$	0.002	0.012	0.033	0.016	0.007	0.022	0.175	0.023	0.027
Eff. $\kappa_n=n^{1/8}$	0.004	0.044	0.050	0.033	0.037	0.073	0.194	0.075	0.096
Eff. $\kappa_n=n^{1/4}$	0.002	0.020	0.037	0.018	0.010	0.033	0.178	0.032	0.041
A.S. $\tau_n=\log(\log(n))$	0.029	0.053	0.038	0.024	0.024	0.030	0.025	0.017	0.061
A.S. $\tau_n=\log(n)$	0.022	0.020	0.023	0.013	0.011	0.018	0.012	0.010	0.024
A.S. $\tau_n=n^{1/8}$	0.026	0.052	0.033	0.022	0.021	0.028	0.022	0.015	0.016
A.S. $\tau_n=n^{1/4}$	0.022	0.035	0.023	0.016	0.012	0.019	0.014	0.011	0.022

also more sensitive to the choice of slackness parameters κ_n and τ_n. As a result, the power comparison of tests in detecting deviations away from $\theta = \theta_K$ is not as conclusive as in Table VIII.

In the results reported in Tables II–IV and VI–VII, the performance of statistics based on the efficient estimator is always worst in specifications for which $\Theta_0(\hat{P}_n) = \emptyset$ in a large number of replications. However, upon finding $\Theta_0(\hat{P}_n) = \emptyset$, it is evident that our asymptotic approximation is inadequate; in fact, the developed statistics cannot even be computed. For completeness, it is therefore also important to examine the performance of these procedures conditional on having found $\Theta_0(\hat{P}_n) \neq \emptyset$. These results are reported in Table X. Surprisingly, the procedures perform well, with our confidence intervals and

TABLE VIII

EMPIRICAL POWER $H_0 : \theta_F \in \Theta_0(P)$ (ON FLAT FACE)—NOMINAL SIZE = 0.05

Procedure	$C = 0.5$					$C = 1$				
	$h=0$	$h=2.5$	$h=5$	$h=7.5$	$h=10$	$h=0$	$h=2.5$	$h=5$	$h=7.5$	$h=10$
				$n = 200$ and $K = 5$						
Eff. $\kappa_n = \log(\log(n))$	0.055	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = \log(n)$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = n^{1/8}$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = n^{1/4}$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
A.S. $\tau_n = \log(\log(n))$	0.040	0.227	0.550	0.837	0.971	0.039	0.231	0.550	0.843	0.970
A.S. $\tau_n = \log(n)$	0.017	0.144	0.483	0.827	0.970	0.019	0.143	0.484	0.830	0.969
A.S. $\tau_n = n^{1/8}$	0.039	0.219	0.536	0.833	0.971	0.039	0.221	0.536	0.839	0.970
A.S. $\tau_n = n^{1/4}$	0.024	0.155	0.488	0.829	0.971	0.026	0.158	0.489	0.833	0.970
				$n = 200$ and $K = 9$						
Eff. $\kappa_n = \log(\log(n))$	0.066	0.293	0.685	0.943	0.996	0.073	0.295	0.686	0.943	0.996
Eff. $\kappa_n = \log(n)$	0.057	0.271	0.672	0.940	0.996	0.067	0.279	0.674	0.941	0.996
Eff. $\kappa_n = n^{1/8}$	0.065	0.290	0.682	0.943	0.996	0.072	0.293	0.683	0.943	0.996
Eff. $\kappa_n = n^{1/4}$	0.058	0.276	0.674	0.941	0.996	0.068	0.282	0.675	0.941	0.996
A.S. $\tau_n = \log(\log(n))$	0.016	0.074	0.225	0.488	0.744	0.015	0.072	0.232	0.494	0.745
A.S. $\tau_n = \log(n)$	0.008	0.045	0.186	0.447	0.728	0.009	0.049	0.190	0.455	0.727
A.S. $\tau_n = n^{1/8}$	0.014	0.071	0.222	0.486	0.743	0.014	0.068	0.229	0.491	0.744
A.S. $\tau_n = n^{1/4}$	0.011	0.059	0.207	0.470	0.735	0.011	0.058	0.209	0.473	0.736
				$n = 1000$ and $K = 5$						
Eff. $\kappa_n = \log(\log(n))$	0.049	0.285	0.702	0.954	0.998	0.049	0.285	0.702	0.954	0.998
Eff. $\kappa_n = \log(n)$	0.048	0.285	0.701	0.954	0.998	0.048	0.285	0.701	0.954	0.998
Eff. $\kappa_n = n^{1/8}$	0.048	0.285	0.702	0.954	0.998	0.048	0.285	0.702	0.954	0.998
Eff. $\kappa_n = n^{1/4}$	0.048	0.285	0.701	0.954	0.998	0.048	0.285	0.701	0.954	0.998
A.S. $\tau_n = \log(\log(n))$	0.050	0.295	0.709	0.952	0.998	0.048	0.294	0.708	0.954	0.997
A.S. $\tau_n = \log(n)$	0.045	0.223	0.566	0.884	0.988	0.043	0.221	0.567	0.884	0.988
A.S. $\tau_n = n^{1/8}$	0.050	0.295	0.709	0.952	0.997	0.048	0.294	0.708	0.954	0.997
A.S. $\tau_n = n^{1/4}$	0.049	0.282	0.645	0.903	0.988	0.048	0.282	0.646	0.904	0.988
				$n = 1000$ and $K = 9$						
Eff. $\kappa_n = \log(\log(n))$	0.054	0.209	0.529	0.851	0.987	0.054	0.209	0.529	0.851	0.987
Eff. $\kappa_n = \log(n)$	0.048	0.193	0.508	0.844	0.985	0.048	0.194	0.508	0.844	0.985
Eff. $\kappa_n = n^{1/8}$	0.054	0.208	0.526	0.850	0.987	0.054	0.208	0.526	0.850	0.987
Eff. $\kappa_n = n^{1/4}$	0.050	0.197	0.509	0.844	0.985	0.050	0.198	0.509	0.844	0.985
A.S. $\tau_n = \log(\log(n))$	0.027	0.109	0.333	0.679	0.926	0.028	0.112	0.332	0.680	0.927
A.S. $\tau_n = \log(n)$	0.011	0.072	0.256	0.600	0.894	0.011	0.071	0.255	0.595	0.892
A.S. $\tau_n = n^{1/8}$	0.023	0.106	0.330	0.676	0.921	0.024	0.107	0.329	0.675	0.922
A.S. $\tau_n = n^{1/4}$	0.014	0.081	0.269	0.604	0.895	0.015	0.082	0.269	0.601	0.894

TABLE IX

EMPIRICAL POWER $H_0: \theta_K \in \Theta_0(P)$ (ON KINK)—NOMINAL SIZE = 0.05

	C = 0.5					C = 1				
Procedure	$h=0$	$h=2.5$	$h=5$	$h=7.5$	$h=10$	$h=0$	$h=2.5$	$h=5$	$h=7.5$	$h=10$
	$n=200$ and $K=5$									
Eff. $\kappa_n=\log(\log(n))$	0.044	0.904	1.000	1.000	1.000	0.062	0.977	1.000	1.000	1.000
Eff. $\kappa_n=\log(n)$	0.005	0.526	0.998	1.000	1.000	0.024	0.921	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/8}$	0.039	0.892	1.000	1.000	1.000	0.057	0.976	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/4}$	0.013	0.734	0.999	1.000	1.000	0.037	0.957	1.000	1.000	1.000
A.S. $\tau_n=\log(\log(n))$	0.035	0.784	1.000	1.000	1.000	0.027	0.896	1.000	1.000	1.000
A.S. $\tau_n=\log(n)$	0.019	0.751	1.000	1.000	1.000	0.023	0.891	1.000	1.000	1.000
A.S. $\tau_n=n^{1/8}$	0.032	0.781	1.000	1.000	1.000	0.026	0.896	1.000	1.000	1.000
A.S. $\tau_n=n^{1/4}$	0.020	0.762	1.000	1.000	1.000	0.025	0.896	1.000	1.000	1.000
	$n=200$ and $K=9$									
Eff. $\kappa_n=\log(\log(n))$	0.056	0.665	0.986	1.000	1.000	0.090	0.895	1.000	1.000	1.000
Eff. $\kappa_n=\log(n)$	0.017	0.346	0.963	0.999	1.000	0.017	0.577	0.995	1.000	1.000
Eff. $\kappa_n=n^{1/8}$	0.050	0.632	0.985	1.000	1.000	0.080	0.872	0.999	1.000	1.000
Eff. $\kappa_n=n^{1/4}$	0.025	0.457	0.976	0.999	1.000	0.035	0.711	0.997	1.000	1.000
A.S. $\tau_n=\log(\log(n))$	0.013	0.322	0.881	0.970	0.983	0.017	0.495	0.939	0.979	0.987
A.S. $\tau_n=\log(n)$	0.010	0.313	0.881	0.970	0.983	0.013	0.481	0.939	0.979	0.987
A.S. $\tau_n=n^{1/8}$	0.012	0.321	0.881	0.970	0.983	0.015	0.494	0.939	0.979	0.987
A.S. $\tau_n=n^{1/4}$	0.010	0.315	0.881	0.970	0.983	0.013	0.490	0.939	0.979	0.987
	$n=1000$ and $K=5$									
Eff. $\kappa_n=\log(\log(n))$	0.050	0.937	1.000	1.000	1.000	0.054	0.961	1.000	1.000	1.000
Eff. $\kappa_n=\log(n)$	0.012	0.811	1.000	1.000	1.000	0.033	0.931	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/8}$	0.044	0.934	1.000	1.000	1.000	0.050	0.960	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/4}$	0.020	0.864	1.000	1.000	1.000	0.037	0.944	1.000	1.000	1.000
A.S. $\tau_n=\log(\log(n))$	0.053	0.917	1.000	1.000	1.000	0.038	0.899	1.000	1.000	1.000
A.S. $\tau_n=\log(n)$	0.020	0.848	1.000	1.000	1.000	0.023	0.899	1.000	1.000	1.000
A.S. $\tau_n=n^{1/8}$	0.052	0.908	1.000	1.000	1.000	0.033	0.899	1.000	1.000	1.000
A.S. $\tau_n=n^{1/4}$	0.035	0.869	1.000	1.000	1.000	0.023	0.899	1.000	1.000	1.000
	$n=1000$ and $K=9$									
Eff. $\kappa_n=\log(\log(n))$	0.047	0.601	0.995	1.000	1.000	0.082	0.944	1.000	1.000	1.000
Eff. $\kappa_n=\log(n)$	0.007	0.303	0.979	1.000	1.000	0.022	0.661	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/8}$	0.037	0.547	0.993	1.000	1.000	0.073	0.935	1.000	1.000	1.000
Eff. $\kappa_n=n^{1/4}$	0.010	0.331	0.983	1.000	1.000	0.033	0.780	1.000	1.000	1.000
A.S. $\tau_n=\log(\log(n))$	0.024	0.532	0.999	1.000	1.000	0.030	0.829	1.000	1.000	1.000
A.S. $\tau_n=\log(n)$	0.011	0.473	0.999	1.000	1.000	0.018	0.803	1.000	1.000	1.000
A.S. $\tau_n=n^{1/8}$	0.021	0.524	0.999	1.000	1.000	0.028	0.823	1.000	1.000	1.000
A.S. $\tau_n=n^{1/4}$	0.012	0.486	0.999	1.000	1.000	0.019	0.803	1.000	1.000	1.000

TABLE X

STATISTICS CONDITIONAL ON $\Theta_0(\hat{P}_n) \neq \emptyset$[a]

Specification	Med. $d_H(\Theta_0(\hat{P}_n), \Theta_0(P))$	$\Theta_0(P)$ CI Coverage	θ_0 on Flat Face Size	θ_0 on Kink Size
		$n = 200$		
$K = 9$, $C = 0.1$	0.200	0.989	0.005	0.016
$K = 15$, $C = 0.1$	0.250	0.998	0.000	0.017
		$n = 500$		
$K = 9$, $C = 0.1$	0.133	0.980	0.010	0.026
$K = 15$, $C = 0.1$	0.202	0.987	0.002	0.027
		$n = 200$		
$K = 9$, $C = 0.1$	0.093	0.978	0.017	0.034
$K = 15$, $C = 0.1$	0.157	0.978	0.003	0.054

[a]Empirical size for tests of $H_0 : \theta_0 \in \Theta_0(P)$ reported for $\kappa_n = \log(\log(n))$.

tests actually being conservative in such instances. We emphasize, however, that there is no reason to expect the results of Table X to hold in generality. Thus, special care should be taken in applying procedures based on the efficient estimator whenever there is reason to doubt the relevance of Assumption 3.6(iv).

REMARK G.1: Since each function $\theta \mapsto F^{(i)}(\int m(x, \theta)\, d\hat{P}_n(x))$ is linear for all $1 \leq i \leq d_F$, the sets $\hat{\Theta}_n(\tau_n)$ are convex polygons. Moreover, their support functions are easily computable through the optimization problem[17]

$$\text{(G.9)} \qquad \nu\big(p, \hat{\Theta}_n(\tau_n)\big) = \sup_{\theta} \langle p, \theta \rangle \quad \text{s.t.}$$

$$F^{(i)}\bigg(\int m(x, \theta)\, d\hat{P}_n(x)\bigg) \leq \frac{\tau_n}{\sqrt{n}} \hat{\sigma}^{(i)} \quad \text{for } i = 1, \dots, d_F.$$

In our simulations, we approximate $\mathbb{S}^2$ by letting $\mathcal{G}$ be a 100 point grid of $[-\pi, \pi]$, and considering the vectors

$$\text{(G.10)} \qquad p(\gamma) \equiv \big(\sin(\gamma), \cos(\gamma)\big)$$

for $\gamma \in \mathcal{G}$. Exploiting (9), we then approximate $d_H(\hat{\Theta}_n(\tau_n), \Theta_0(P))$ by $\max_{\gamma \in \mathcal{G}} |\nu(p(\gamma), \hat{\Theta}_n(\tau_n)) - \nu(p(\gamma), \Theta_0(P))|$.

[17]This problem is easily solvable by standard packages. We employ the open software Matlab toolboxes YALMIP and MPT, available at http://users.isy.liu.se/johanl/yalmip/ and http://control.ee.ethz.ch/~mpt/.

REMARK G.2: Because, in this context, all constraints are linear in θ, the support function has the dual representation

$$\text{(G.11)}\quad \nu\big(p, \Theta_0(\hat{P}_n)\big) = \min_{w \in \mathbf{R}_+^{d_F}} \left\langle w, F_S\left(\int m_S(x,\theta)\, d\hat{P}_n(x)\right)\right\rangle \quad \text{s.t.} \quad A'w = p,$$

where A and $v \mapsto F_S(v)$ are as defined in Example 2.1, and $m_S(x,\theta)$ is constant in $\theta \in \Theta$ (see (F.3)). Moreover, the minimizers of (G.11) are the Lagrange multipliers $\lambda(p, \hat{P}_n)$ of the primal problem that defines $\nu(p, \Theta_0(\hat{P}_n))$. Therefore, by (23) and direct calculation, solving (G.11) suffices for computing the bootstrap process G_n^* given by

$$\text{(G.12)}\quad G_n^*(p) = -\lambda(p, \hat{P}_n)' \nabla F_S\left(\frac{1}{n}\sum_{i=1}^n m_S(X_i,\theta)\right) \times \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\left\{m_S(X_i,\theta) - \frac{1}{n}\sum_{i=1}^n m_S(X_i,\theta)\right\}.$$

In our simulations, we draw W_i from the Rademacher distribution, that is, $P(W_i = 1) = P(W_i = -1) = 1/2$, and we compute the critical value $\hat{c}_{1-\alpha}$ as the $1-\alpha$ quantile across bootstrap replications of

$$\text{(G.13)}\quad \sup_{\gamma \in \mathcal{G}} \max\big\{G_n^*\big(p(\gamma)\big), 0\big\},$$

where $p(\gamma)$ and $\mathcal{G}$ are as in (G.10). The support function for the confidence region $\hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}$ (as in Example 5.1) is then given by $\nu(\cdot, \hat{\Theta}_n) + \hat{c}_{1-\alpha}/\sqrt{n}$, and hence we check whether $\Theta_0(P) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}$ by verifying that $\nu(p(\gamma), \Theta_0(P)) \leq \nu(p(\gamma), \hat{\Theta}_n) + \hat{c}_{1-\alpha}/\sqrt{n}$ for all $\gamma \in \mathcal{G}$; see also Beresteanu and Molinari (2008).

REMARK G.3: In order to compute $\hat{c}^B_{1-\alpha}(\tau_n)$, we draw samples $\{X_i^*\}_{i=1}^n$ from $\{X_i\}_{i=1}^n$ with replacement, let $\hat{P}_n^*$ denote the empirical measure induced by $\{X_i^*\}_{i=1}^n$, and let $(\hat{\sigma}_n^{*(i)})^2$ be the corresponding estimate of the asymptotic variance of constraint number i. We then obtain $\hat{c}^B_{1-\alpha}(\tau_n)$ by computing the $1-\alpha$ quantile across bootstrap replications of

$$\text{(G.14)}\quad \sup_{\theta \in \hat{\Theta}_n(\tau_n)} \max_{1 \leq i \leq d_F} \left\{\sqrt{n}\left(\frac{1}{\hat{\sigma}_n^{*(i)}} F^{(i)}\left(\int m(x,\theta)\, d\hat{P}_n^*(x)\right) - \frac{1}{\hat{\sigma}_n^{(i)}} F^{(i)}\left(\int m(x,\theta)\, d\hat{P}_n(x)\right)\right)_+ \times \omega_n^{(i)}(\theta)\right\},$$

where $\omega_n^{(i)}(\theta) \equiv 1\{|F^{(i)}(\int m(x,\theta)\, d\hat{P}_n(x))| \leq \tau_n \hat{\sigma}_n^{(i)}/\sqrt{n}\}$. Since $\mathrm{CS}(\tau_n)$ is a convex polygon, we compute its support function in a manner analogous to

(G.9), and check whether $\Theta_0(P) \subseteq \mathrm{CS}(\tau_n)$ by verifying that $\nu(p(\gamma), \Theta_0(P)) \leq \nu(p(\gamma), \mathrm{CS}(\tau_n))$ for all $\gamma \in \mathcal{G}$, where $p(\gamma)$ and $\mathcal{G}$ are as in (G.10).

REMARK G.4: Following the construction of $\hat{c}^B_{1-\alpha}(\tau_n)$, to obtain $\hat{c}^{AS}_{1-\alpha}(\theta)$ we draw samples $\{X_i^*\}_{i=1}^n$ from $\{X_i\}_{i=1}^n$ with replacement, let $\hat{P}_n^*$ denote the empirical measure induced by $\{X_i^*\}_{i=1}^n$, and let $(\hat{\sigma}_n^{*(i)})^2$ be the corresponding estimate of the asymptotic variance of constraint i. For $\omega_n^{(i)}(\theta) \equiv 1\{|F^{(i)}(\int m(x,\theta)\, d\hat{P}_n(x))| \leq \tau_n \hat{\sigma}_n^{(i)}/\sqrt{n}\}$ and

$$
Q_n^*(\theta) \equiv \max_{1\leq i\leq d_F}\left\{\left(\frac{1}{\hat{\sigma}_n^{*(i)}}F^{(i)}\left(\int m(x,\theta)\, d\hat{P}_n^*(x)\right) - \frac{1}{\hat{\sigma}_n^{(i)}}F^{(i)}\left(\int m(x,\theta)\, d\hat{P}_n(x)\right)\right)_+ \times \omega_n^{(i)}(\theta)\right\}, \tag{G.15}
$$

we then let $\hat{c}^{AS}_{1-\alpha}(\theta)$ be the $1-\alpha$ quantile of $\sqrt{n}Q_n^*(\theta)$ across 200 bootstrap replications.

REFERENCES

ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis—A Hitchhiker's Guide*. Berlin: Springer-Verlag. [8,9,14,20,21,24,45,54,66]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. [79,84,85]

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [82,91]

BHATIA, R. (1997): *Matrix Analysis*. New York: Springer-Verlag. [13,60]

BICKEL, P. J., C. A. KLASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer. [1,18,20,21,37,45,46,57,59, 66]

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155. [61,67]

BUGNI, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753. [79,80,82]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [80,82]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. Providence: American Mathematical Society. [54]

DEMBO, A., AND O. ZEITOUNI (1998): *Large Deviations Techniques and Applications*. New York: Springer. [26]

HO, K. (2009): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 99, 1039–1079. [72]

KAIDO, H. (2012): "A Dual Approach to Inference for Partially Identified Econometric Models," Working Paper, Boston University. [54,55]

KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer-Verlag. [37,49,53,55,56]

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. New York: Springer. [57,58]

LI, S., Y. OGURA, AND V. KREINOVICH (2002): *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables (Theory and Decision Library B)*. Berlin: Springer-Verlag. [40,54]

LUENBERGER, D. G. (1969): *Optimization by Vector Space Methods*. New York: Wiley. [5,11,14, 42]

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601. [15,24,46]

ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton: Princeton University Press. [6,11,44]

VAN DER VAART, A. W. (1991): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178–204. [55]

——— (1999): *Asymptotic Statistics*. New York: Cambridge University Press. [56,58]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer-Verlag. [23,36,37,39,49,51-53]

Dept. of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, U.S.A.; hkaido@bu.edu

and

Dept. of Economics, University of California—San Diego, 9500 Gilman Drive MS 0508, La Jolla, CA 92093-0508, U.S.A.; a2santos@ucsd.edu.

Manuscript received May, 2011; final revision received June, 2013.