ONLINE APPENDIX FOR: "INFORMATION BASED INFERENCE IN MODELS WITH SET-VALUED PREDICTIONS AND MISSPECIFICATION"

APPENDIX B: LEMMAS USED IN PROOFS OF MAIN THEOREMS

B.1. *Pseudo-True Sets and Score Function*

We give three lemmas that we use to show the differentiability of $L(\theta|x)$. To ease notation, we drop conditioning on X. We let $\mathcal{C}$ denote the collection of closed subsets of $\mathcal{Y}$. Molchanov and Molinari (2018, Section 2.2) show that $\operatorname{core}(\nu_\theta(\cdot))$ can be expressed as $\operatorname{core}(\nu_\theta(\cdot)) \equiv \{Q \in \mathcal{M}(\Sigma_Y) : \nu_\theta(A) \le Q(A) \le \nu_\theta^*(A), A \subseteq \mathcal{C}\}$, with $\nu_\theta^*(A) \equiv \int_{\mathcal{U}} \mathbf{1}(G(u;\theta) \cap A \neq \emptyset) dF_\theta(u)$.[1] Let $\mathcal{A} \subseteq 2^{\mathcal{Y}}$ be a collection of events. Among sets in $\mathcal{A}$, let $\mathcal{A}_=$ collect all restrictions such that $\nu_\theta(A) = \nu_\theta^*(A)$. That is, the sets belonging to $\mathcal{A}_=$ imply equality restrictions. We then let $\mathcal{A}_\ge$ collect the remaining events. Let Δ denote the $|\mathcal{Y}| - 1$ dimensional unit-simplex. Consider the following problem:

$$\mathbf{P}(\mathcal{A}): \qquad \upsilon(\theta;\mathcal{A}) \equiv \max_{q \in \Delta} \sum_{y \in \mathcal{Y}} p_0(y) \ln q(y) \tag{B.1}$$

$$s.t. \quad \sum_{y \in A} q(y) \ge \nu_\theta(A), \;\; A \in \mathcal{A}_\ge \tag{B.2}$$

$$\sum_{y \in A} q(y) = \nu_\theta(A), \;\; A \in \mathcal{A}_=, \tag{B.3}$$

Let $A \subset \mathcal{Y}$. As $\mathcal{Y}$ is finite, one may represent the probability that any distribution P with probability mass function $p \in \Delta$ assigns to a set A through a *representer* a of A, by writing

$$P(A) = p^\top a,$$

with $a \in \{0,1\}^{|\mathcal{Y}|}$. For example, take $\mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$, $A = \{(1,0),(1,1)\}$, and $a = (0,0,1,1)^\top$. Then, $P(A) = p^\top a$. Similarly, for some $b_A \in \{0,1\}^{|\mathcal{Y}|}$, the constraints

[1] For a given compact set $A \subseteq \mathcal{Y}$, $\nu_\theta^*(A)$ is the *capacity functional* of $G(u;\theta)$ (Molchanov and Molinari, 2018, Definition 1.23), and $\nu_\theta(A) = 1 - \nu_\theta^*(A^c)$, with A^c the complement of A and $\nu_\theta^*(\cdot)$ extended to the family of open sets (e.g., Molchanov and Molinari, 2018, p.20), and where ν_θ is the containment functional of $G(\cdot;\theta)$.

in (B.2)-(B.3) can be written as

$$q^\top b_A \geq \nu_\theta(A),\ A \in \mathcal{A}_\geq \qquad \text{and} \qquad q^\top b_A = \nu_\theta(A),\ A \in \mathcal{A}_=.$$

For any pair of sets $A_1, A_2 \subset \mathcal{Y}$ with associated representer vectors $a^1, a^2 \in \{0,1\}^{|\mathcal{Y}|}$, their union $A_1 \cup A_2$ and intersections $A_1 \cap A_2$ are represented by $a^1 \vee a^2$ (componentwise maximum) and $a^1 \wedge a^2$ (componentwise minimum). For any event $A \subseteq \mathcal{Y}$ with $A = \cup_{i=1}^k A_i$,

$$P(A) = \sum_{I \neq \emptyset, I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} P\big(\bigcap_{i \in I} A_i\big).$$

In terms of corresponding vectors, $p^\top a = \sum_{I \neq \emptyset, I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} p^\top (\bigwedge_{i \in I} a^i)$. Since this holds for any p in the probability simplex, we must have

$$a = \sum_{I \neq \emptyset, I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} (\bigwedge_{i \in I} a^i). \tag{B.4}$$

This means that the representers of the events A and $A_1, \dots, A_k$ are linearly dependent. The following lemma shows that the opposite is also true.

LEMMA B.1: *Let $a^0, \dots, a^k$ be the representers of $A_0, \dots, A_k$. The following statements are equivalent: (1) $\big\{ \bigwedge_{i \in I} a^i, I \neq \emptyset, I \subseteq \{0, \dots, k\} \big\}$ are linearly dependent; (2) There exists $j \in \{0, \dots, k\}$ such that $A_j = \bigcup_{i \in \{0,\dots,k\} \setminus \{j\}} A_i$.*

PROOF: Order the elements of $\mathcal{Y}$ as $\{y_1, y_2, \dots, y_{|\mathcal{Y}|}\}$. (2. $\Rightarrow$ 1.) follows from (B.4). For (1. $\Rightarrow$ 2.), w.l.o.g. take $j = 0$ and suppose $C \equiv A_0 \setminus \cup_{i=1}^k A_i$ is nonempty. Collect the indices of outcomes belonging to C in $I_C \subset \{1, \dots, |\mathcal{Y}|\}$. Take $y_j \in C \subset \mathcal{Y}$. The representer e_j of y_j is a $|\mathcal{Y}|$-dimensional vector whose j-th component is 1, and the remaining components are all 0s. Note that $y_j \notin \cup_{i=1}^k A_i$ implies that the j-th component of $\bigwedge_{i \in I} a^i$ is 0 for any $I \neq \emptyset, I \subseteq \{1, \dots, k\}$. Hence, e_j cannot be expressed as a linear combination of $\big\{ \bigwedge_{i \in I} a^i, I \neq \emptyset, I \subseteq \{1, \dots, k\} \big\}$. As $y_j \in A_0$, this means a^0 cannot be expressed as linear combination of $\big\{ \bigwedge_{i \in I} a^i, I \neq \emptyset, I \subseteq \{1, \dots, k\} \big\}$. Hence, $\big\{ a^0, \bigwedge_{i \in I} a^i, I \neq \emptyset, I \subseteq \{1, \dots, k\} \big\}$ are linearly independent. The case with $\cup_{i=1}^k A_i \setminus A_0 \neq \emptyset$ can be analyzed similarly. *Q.E.D.*

LEMMA B.2: *Let Assumption 1 hold. Then (i) there exists a collection* $\mathcal{A}^{(*e)} \subseteq 2^{\mathcal{Y}}$ *not dependent on* $\theta \in \Theta$ *such that* $\mathrm{core}(\nu_\theta(\cdot)) = \{Q \in \mathcal{M}(\Sigma_Y) : Q(A) \geq \nu_\theta(A), A \in \mathcal{A}^{(*e)}\}$ *for* $\mathrm{core}(\nu_\theta(\cdot))$ *in* (3.2)*, and no collection* $\mathcal{A}^* \subseteq 2^{\mathcal{Y}}$ *of cardinality smaller than* $\mathcal{A}^{(*e)}$ *suffices to characterize* $\mathrm{core}(\nu_\theta(\cdot))$*; (ii) the optimal value of* $\mathbf{P}(\mathcal{A}^{(*e)})$ *is* $L(\theta|x)$*; (iii) the solution* $q^* \in \Delta$ *to problem* $\mathbf{P}(\mathcal{A}^{(*e)})$ *is unique and also solves problem* $\mathbf{P}(\mathcal{A}^{all})$ *with* $\mathcal{A}^{all} = \{A : A \subseteq \mathcal{Y},\ A \text{ closed}\}$*; (iv) its associated Lagrange multiplier vector* λ^* *is unique.*

PROOF: **Part (i).** A family of closed sets $\mathcal{A}^*$ is a *core determining class* (Galichon and Henry, 2011) if any probability measure Q defined on $\mathcal{Y}$ satisfying the inequalities $Q(A) \geq \nu_\theta(A)$ for all $A \in \mathcal{A}^*$ satisfies the inequalities $Q(A) \geq \nu_\theta(A)$ for all closed sets $A \subseteq \mathcal{Y}$. A family of closed sets $\mathcal{A}$ is a *smallest* core determining class (Luo, Ponomarev, and Wang, 2025) if it has the smallest cardinality among all core determining classes. Part (i) follows directly by Corollary 1.1 in Luo, Ponomarev, and Wang (2025), due to Assumption 1-b, as their result shows that if the support of the random set $G(U|x;\theta)$, conditional on $X = x$ does not depend on θ, neither does the smallest core determining class.

Part (ii). By the definition of a smallest core determining class, the collection of inequalities in $\mathcal{A}^{(*e)}$ yields the same constraint set as in (3.23). The solution to $\mathbf{P}(\mathcal{A}^{(*e)})$ exists by the continuity of the objective function and the compactness of the probability simplex.

Part (iii). As $q \mapsto \mathbb{E} \ln q$ is strictly concave and the domain of q is convex, uniqueness of q^* follows. As the collection of inequalities in $\mathcal{A}^{(*e)}$ yields the same constraint set as in (3.23), q^* solves also the original problem in (3.23).

Part (iv). The constraint set consists of linear (in)equalities. Hence, the Karush-Kuhn-Tucker conditions hold at the feasible point q^* with Lagrange multiplier λ^*. A sufficient condition for λ^* to be unique is that the Linear Independence Constraint Qualification (LICQ) holds. To establish this, we first note that the full set of constraints can be expressed as (e.g., Molchanov and Molinari, 2018, Section 2.2):

$$\nu_\theta^*(A) \geq q^\top b_A \geq \nu_\theta(A),\ A \subset \mathcal{Y} : 1 \leq |A| \leq \lceil |\mathcal{Y}|/2 \rceil, \tag{B.5}$$

where $|\cdot|$ denotes the cardinality of the set in its argument, and $\lceil \cdot \rceil$ represents the smallest integer greater than or equal to its argument. (B.5) follows because for any set $A \subset \mathcal{Y}$ and its complement $A^c = \mathcal{Y} \setminus A$, one has $b_A = 1 - b_{A^c}$ and $\nu_\theta(A) = 1 - \nu_\theta^*(A^c)$. This implies that the gradient of any pair of inequalities in (B.5) equals a representer b_A. Moreover, either only one of the two inequalities in (B.5) can be an *active* inequality, or we are in the presence of an equality restriction. Next, recall that we are further restricting the collection of sets in (B.5) to be the ones in $\mathcal{A}^{(*e)}$. If the LICQ condition fails, there must exist a collection of sets $A_j \in \{A \subset \mathcal{A}^{(*e)} : 1 \leq |A| \leq \lceil |\mathcal{Y}|/2 \rceil\}$, $j = 1, \ldots, k$, such that $q^\top b_{A_j} = \nu_\theta(A_j)$, $q^\top b_A = \nu_\theta(A)$, and, by Lemma B.1, $A = \cup_{j=1,\ldots,k} A_j$. This in turn implies

$$\nu_\theta(A) \geq \sum_{I \neq \emptyset, I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \nu_\theta\Big(\bigcap_{i \in I} A_i\Big),$$

and we have that the inequality for the set A is satisfied whenever the inequalities for the sets $A_j, j = 1, \ldots, k$ are satisfied. But this contradicts $\mathcal{A}^{(*e)}$ being an exact core determining class because such a set A could be removed from it. *Q.E.D.*

LEMMA B.3: *Under the assumptions of Theorem 3.2, $\phi(\cdot,\cdot)$ as defined in (3.28) is jointly continuous in (θ, p).*

PROOF: Passing the infimum in (3.28) inside the integral and recalling $\mathfrak{q}_{\vartheta,x}$ from (3.5), we can write $\phi(\theta, p) = \int_{\mathcal{X}} p_x(x) v(\theta, p_{y|x}, x) d\xi(x)$, where $v(\theta, p_{y|x}; x) = \inf_{q_{y|x} \in \mathfrak{q}_{\vartheta,x}} I(p_{y|x} || q_{y|x})$ and, denoting $H(p_{y|x}) \equiv -\int_{\mathcal{Y}} p_{y|x}(y|x) \ln p_{y|x}(y|x) d\mu(y)$,

$$I(p_{y|x}||q_{y|x}) = -H(p_{y|x}) - \int_{\mathcal{Y}} p_{y|x}(y|x) \ln q_{y|x}(y|x) d\mu(y). \tag{B.6}$$

Let (θ_n, p_n) be a sequence such that $(\theta_n, p_n) \to (\theta, p)$, where $p_{n,y|x}(y|x) - p_{y|x}(y|x) \to 0$ uniformly in y and x, and $\|p_{n,x} - p_x\|_{L^1_\xi} \to 0$. Then,

$$\phi(\theta_n, p_n) - \phi(\theta, p) = \int_{\mathcal{X}} p_{n,x}(x) v(\theta_n, p_{n,y|x}, x) d\xi(x) - \int_{\mathcal{X}} p_x(x) v(\theta, p_{y|x}, x) d\xi(x)$$

$$= \underbrace{\int_{\mathcal{X}} p_{n,x}(x)[v(\theta_n, p_{n,y|x}, x) - v(\theta, p_{y|x}, x)] d\xi(x)}_{(i)} + \underbrace{\int_{\mathcal{X}} [p_{n,x}(x) - p_x(x)] v(\theta, p_{y|x}, x) d\xi(x)}_{(ii)}.$$

Using Assumption 1-(e), $|v(\theta, p_{y|x}, x)| \le \sum_{y\in\mathcal{Y}} p_{y|x}(y|x)\ln\frac{1}{c} \le d_Y \ln\frac{1}{c} =: K$, and (ii) can be bounded by $K\|p_{n,x}-p_x\|_{L^1_\xi} \to 0$. Term (i) can be further decomposed as follows:

$$\underbrace{\int_{\mathcal{X}} p_{n,x}(x)[v(\theta_n, p_{n,y|x}, x) - v(\theta, p_{n,y|x}, x)]d\xi(x)}_{(i-a)} + \underbrace{\int_{\mathcal{X}} p_{n,x}(x)[v(\theta, p_{n,y|x}, x) - v(\theta, p_{y|x}, x)]d\xi(x)}_{(i-b)}.$$

Recall that $v(\theta_n, p_{n,y|x}, x) = -H(p_{n,y|x}) - \mathbb{E}_{p_{n,y|x}}[\ln q^*_{\theta_n,y|x}(Y|x)|x]$. By the mean-value theorem and Theorem 3.1, there exists $\tilde\theta_n$ between θ_n and θ such that

$$(i-a) \le \int_{\mathcal{X}} p_{n,x}(x)\left|\mathbb{E}_{p_{n,y|x}}[s_{\tilde\theta_n,y|x}(Y|x;p_{n,y|x})|x]\right|\|\theta_n-\theta\|d\xi(x) \le M^{1/2}\|\theta_n-\theta\|,$$

where the last inequality follows because $\sup_{\theta\in\Theta}\mathbb{E}[\|s_\theta(Y|X;p_{y|x})\|^2] \le M$ and

$$\|E_{p_{n,y|x}}[s_{\tilde\theta_n,y|x}(Y|X;p_{n,y|x})|X]\|_{L^1_{p_{n,y|x}}} \le \|E_{p_{n,y|x}}[s_{\tilde\theta_n,y|x}(Y|X;p_{n,y|x})|X]\|_{L^2_{p_{n,y|x}}} \le M^{1/2}.$$

Hence, $(i-a)$ tends to 0 as $\theta_n \to \theta$. To show that also $(i-b)$ vanishes, let $\|p_{y|x}-p'_{y|x}\|_1 = \sum_{y\in\mathcal{Y}}|p_{y|x}(y|x)-p'_{y|x}(y|x)|$. For any θ and $p_{y|x}, p'_{y|x}$ such that $\|p_{y|x}-p'_{y|x}\|_1 \le 1/2$,

$$\begin{aligned}|v(\theta,p_{y|x};x)-v(\theta,p'_{y|x};x)| &\le \left|\inf_{q_{y|x}\in\mathfrak{q}_{\vartheta,x}} I(p_{y|x}\|q_{y|x}) - \inf_{q_{y|x}\in\mathfrak{q}_{\vartheta,x}} I(p'_{y|x}\|q_{y|x})\right| \\ &\le \sup_{q_{y|x}\in\mathfrak{q}_{\vartheta,x}} |I(p_{y|x}\|q_{y|x}) - I(p'_{y|x}\|q_{y|x})| \le \omega(\|p_{y|x}-p'_{y|x}\|_1),\end{aligned}$$

where $\omega(v) = v\ln\frac{d_Y}{v} + v\ln\frac{1}{c}$ by Lemma 2.7 in Csiszár and Körner (2011) and (B.6), and the convergence is uniform in x as we assumed $p_{n,y|x}(y|x) - p_{y|x}(y|x) \to 0$ uniformly in y and x. *Q.E.D.*

B.2. *Derivation of Results and Verification of Conditions for the Entry Game Example 1*

PROPOSITION B.1: *Under the assumptions laid out in Example 1, (i) the profiled likelihood $q^*_{\theta,y|x}$ for $y\in\{(0,0),(0,1),(1,0),(1,1)\}$ is given in equations (3.17)-(3.20). (ii) The score function is given in equations (3.24)-(3.27).*

PROOF: In this example $q_{y|x}((0,1)|x) = \eta_1(\theta;x) - q_{y|x}((1,0)|x)$. Let $z \equiv q_{y|x}((1,0)|x)$ and $c(\theta) = p_{0,y|x}((0,0)|x)\ln F_\theta(S_{\{(0,0)\}|x;\theta}) + p_{0,y|x}((1,1)|x)\ln F_\theta(S_{\{(1,1)\}|x;\theta})$. Using that $\sum_{\tilde{y}} q_{y|x}(\tilde{y}|x) = 1$, rewrite the optimization problem as

$$V(\theta) = \sup_{z} c(\theta) + p_{0,y|x}((1,0)|x)\ln z + p_{0,y|x}((0,1)|x)\ln(\eta_1(\theta;x) - z)$$
$$s.t.\ z - \eta_3(\theta;x) \geq 0 \quad \text{and} \quad \eta_2(\theta;x) - z \geq 0,$$

Define the Lagrangian of this problem by

$$\mathcal{L}(z,\lambda,\theta) = c(\theta) + p_{0,y|x}((1,0)|x)\ln z + p_{0,y|x}((0,1)|x)\ln(\eta_1(\theta;x) - z)$$
$$+ \lambda_1(z - \eta_3(\theta;x)) + \lambda_2(\eta_2(\theta;x) - z).$$

Part (i). Since $c(\theta)$ does not affect the solution, we drop it in what follows. The Karush-Kuhn-Tucker (KKT) conditions of this problem are $-p_{0,y|x}(1,0|x)\frac{1}{z} + p_{0,y|x}(0,1|x)\frac{1}{\eta_1(\theta;x)-z} - \lambda_1 + \lambda_2 = 0$, $\lambda_1(\eta_3(\theta;x) - z) = 0$, $\lambda_2(z - \eta_2(\theta;x)) = 0$, and $\lambda_1, \lambda_2 \geq 0$. Then, (3.17)-(3.20) are obtained by solving the KKT conditions for three cases: (1) $\lambda_1 = \lambda_2 = 0$; (2) $\lambda_2 > 0$; (3) $\lambda_1 > 0$. (See Kaido and Molinari, 2024, pp.6-7, for a full derivation.)

Part (ii). To establish this result, we let $\theta(t) = \theta + th, h \in \mathbb{R}^d$ and define $\tilde{V}(t) = V(\theta(t)),\ \tilde{\mathcal{L}}(z,\lambda,t) = \mathcal{L}(z,\lambda,\theta(t))$, so that the Lagrangian becomes

$$\tilde{\mathcal{L}}_t(z,\lambda,\theta(t)) = p_{0,y|x}((0,0)|x)\frac{\nabla_\theta F_\theta(S_{\{(0,0)\}|x;\theta})^\top h}{F_\theta(S_{\{(0,0)\}|x;\theta})} + p_{0,y|x}((1,1)|x)\frac{\nabla_\theta F_\theta(S_{\{(1,1)\}|x;\theta})^\top h}{F_\theta(S_{\{(1,1)\}|x;\theta})}$$
$$+ p_{0,y|x}((0,1)|x)\frac{\nabla_\theta \eta_1(\theta;x)^\top h}{\eta_1(\theta;x)-z} + \lambda_1 \nabla_\theta \eta_3(\theta;x)^\top h + \lambda_2 \nabla_\theta \eta_2(\theta;x)^\top h.$$

Hence, for any sequence $\{t_n\}$ with $t_n \downarrow 0$, the maximizer of $\mathcal{L}(z,\lambda,\theta + t_n h)$ exists. The domain of the control variable and parameter $[0,1] \times (-\epsilon,\epsilon)$ is bounded, hence the inf-boundedness assumption of Rockafellar (1984) holds, ensuring that the parametric optimization problem indexed by t is directionally stable in the sense of Gauvin and Janin (1990). Using that (q^*, λ^*) are unique as shown above, we apply Gauvin and Janin (1990, Corollary 4.2) to obtain full differentiability of V. The derivative of $\tilde{V}(t)$ can be obtained analytically solving three cases: (1) $\lambda_1 = \lambda_2 = 0$; (2) $\lambda_1 = 0, \lambda_2 > 0$; (3) $\lambda_1 > 0, \lambda_2 = 0$. (See Kaido and Molinari, 2024, p.8, for a full derivation.) *Q.E.D.*

PROPOSITION B.2: *The mapping $\gamma \mapsto \Theta^*(\gamma)$ in Example 1 on p. 17 is continuous.*

PROOF: Denoting $\vartheta_j = 0.5 + \theta_j$, we have $\Theta^*(p_\gamma) = \Xi^*(p_\gamma) - [0.5\ 0.5]^\top$ for $\Xi^*(p_\gamma) \equiv \{\vartheta \in [0.05, 0.5]^2 : \vartheta_1\vartheta_2 = p_\gamma((1,1)|1);\ \frac{p_\gamma((1,1)|1)}{1-p_\gamma((1,0)|1)} \le \vartheta_1 \le 1 - p_\gamma((0,1)|1)\}$. We then have that for any $\theta \in \Theta$ and $\vartheta = [0.5\ 0.5]^\top + \theta$,

$$\mathrm{dist}(\theta, \Theta^*(p_\gamma)) = \mathrm{dist}(\vartheta, \Xi^*(p_\gamma)) = \min_{\tilde{\vartheta}_1 \in B(p_\gamma)} \left\| \left[\tilde{\vartheta}_1 \ \ \frac{p_\gamma((1,1)|1)}{\tilde{\vartheta}_1}\right]^\top - \vartheta \right\|, \tag{B.7}$$

with $B(p_\gamma) = \left[\max\left\{0.05, 2p_\gamma((1,1)|1), \frac{p_\gamma((1,1)|1)}{1-p_\gamma((1,0)|1)}\right\}, \min\left\{0.5, \frac{p_\gamma((1,1)|1)}{0.05}, 1-p_\gamma((0,1)|1)\right\}\right]$. Recall that $p_\gamma(y|1)$ is an affine function of γ for all $y \in \mathcal{Y}$. Hence, the objective function in the minimization problem in (B.7) is jointly continuous in ϑ_1 and p_γ, and $B(p_\gamma)$ converges in Hausdorff distance to $B(p_0)$ when $p_\gamma \to p_{\gamma_0}$ (where for non-empty intervals $A = [a_1, a_2], B = [b_1, b_2] \subset \mathbb{R}$, $\mathrm{dist}_H(A, B) = \max\{|a_1 - b_1|, |a_2 - b_2|\}$), and hence it is a continuous correspondence on Γ. It follows that all assumptions of Berge's maximum theorem are satisfied, and $\gamma \mapsto \mathrm{dist}(\theta, \Theta^*(p_\gamma))$ is continuous on Γ. Hence, by Rockafellar and Wets (2005, Proposition 5.11), $\gamma \mapsto \Theta^*(p_\gamma)$ is both upper and lower hemicontinuous on Γ, and hence is continuous at $\gamma = 0$. *Q.E.D.*

B.2.1. *Verification of Assumptions*

Assumptions 2, 3 (iv), and the requirement in (3.35), are high-level conditions to be verified in each application of our method. Below we do so for the entry game example, under some regularity conditions. We first provide some notation that will be useful throughout. Let $\|p_{y|x}\|_{\mathcal{H}} = \sup_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} |p_{y|x}(y|x)|$. For $j = 1, 2$, we let:

$$Z_j(X; p_{y|x}) \equiv p_{y|x}((1,0)|X)\eta_1(\theta; X) - (p_{y|x}((1,0)|X) + p_{y|x}((0,1)|X))\eta_{j+1}(\theta; X), \tag{B.8}$$

with the functions $\eta_1(\theta; x), \eta_2(\theta; x), \eta_3(\theta; x)$ defined in (3.11)-(3.13). As $\eta_2(\theta; X) \ge \eta_3(\theta; X)$, $Z_1(x; p_{y|x}) \le Z_2(x; p_{y|x})$. We use the functions Z_1, Z_2 to define indicators $\mathbb{I}_\ell$ that re-express the sets $\Theta_\ell, \ell = 1, 2, 3$, in (3.14)-(3.16):

$$\mathbb{I}_1(x; p_{y|x}) = 1\{Z_1(x; p_{y|x}) \le 0\}1\{Z_2(x; p_{y|x}) \ge 0\} \tag{B.9}$$

$$\mathbb{I}_2(x;p_{y|x}) = 1\{Z_1(x;p_{y|x}) > 0\} \tag{B.10}$$

$$\mathbb{I}_3(x;p_{y|x}) = 1\{Z_2(x;p_{y|x}) < 0\}. \tag{B.11}$$

One may rewrite the score functions in (3.24)-(3.27) as

$$s_\theta((0,0)|x;p_{y|x}) = \frac{\nabla_\theta F_\theta(S_{\{(0,0)\}|x;\theta})}{F_\theta(S_{\{(0,0)\}|x;\theta})},$$
$$s_\theta((1,1)|x;p_{y|x}) = \frac{\nabla_\theta F_\theta(S_{\{(1,1)\}|x;\theta})}{F_\theta(S_{\{(1,1)\}|x;\theta})},$$
$$s_\theta((0,1)|x;p_{y|x}) = \frac{\nabla_\theta \eta_1(\theta;x)}{\eta_1(\theta;x)}\mathbb{I}_1(x;p_{y|x}) + \frac{\nabla_\theta[\eta_1(\theta;x)-\eta_2(\theta;x)]}{\eta_1(\theta;x)-\eta_2(\theta;x)}\mathbb{I}_2(x;p_{y|x}) + \frac{\nabla_\theta[\eta_1(\theta;x)-\eta_3(\theta;x)]}{\eta_1(\theta;x)-\eta_3(\theta;x)}\mathbb{I}_3(x;p_{y|x}),$$
$$s_\theta((1,0)|x;p_{y|x}) = \frac{\nabla_\theta \eta_1(\theta;x)}{\eta_1(\theta;x)}\mathbb{I}_1(x;p_{y|x}) + \frac{\nabla_\theta \eta_2(\theta;x)}{\eta_2(\theta;x)}\mathbb{I}_2(x;p_{y|x}) + \frac{\nabla_\theta \eta_3(\theta;x)}{\eta_3(\theta;x)}\mathbb{I}_3(x;p_{y|x}).$$

For any vector $a = (a_1, \ldots, a_{d_X})$, define the differential operator by $D^{|a|} = \frac{\partial^{|a|}}{\partial x_1^{a_1} \cdots \partial x_d^{a_{d_X}}}$, where $|a| = \sum_i^{d_X} a_i$. Then, for a function $h : \mathcal{X} \to \mathbb{R}$, let

$$\|h\|_{\infty,\alpha} = \max_{|a|\le[\alpha]} \sup_x |D^a h(x)| + \max_{|a|=[\alpha]} \sup_{x\neq x'} \frac{|D^a h(x) - D^a h(x')|}{\|x-x'\|^{\alpha-[\alpha]}}.$$

Let $\mathcal{C}_M^\alpha(\mathcal{X})$ be the set of continuous functions $h : \mathcal{X} \to \mathbb{R}$ with $\|h\|_{\infty,\alpha} \le M$. We next provide regularity conditions under which we verify Assumptions 2 and 3 (iv) .

ASSUMPTION B.1: *For the entry game model in Example 1,*

(i) There exists $C > 0$ s.t. $\|\nabla_\theta \eta_j(\theta;x)\| \le C,\ j = 1, \ldots, 3$, for all $x \in \mathcal{X}$.

(ii) There exists $c > 0$ s.t. $\mathcal{H} = \{p_{y|x} : \mathcal{X} \to [0,1]^{\mathcal{Y}} : p_{y|x}(y|x) \ge c, \forall (y,x) \in \mathcal{Y} \times \mathcal{X}\}$.

(iii) If X has a component with continuous distribution, the probability density function (p.d.f.) of $Z_j|X_d$, for $j = 1, 2$, is uniformly bounded on the support of $Z_j|X_d$, where X_d denotes the subvector of X containing discrete covariates with finite support. If there are no discrete covariates, the restriction is on the unconditional p.d.f. of Z_j.

ASSUMPTION B.2: *(a) $\mathbb{E}\left[\left\|\frac{\nabla_\theta F_\theta(S_{\{y\}|x;\theta})}{F_\theta(S_{\{y\}|x;\theta})}\right\|^2\right] \le C$ for $y = (0,0), (1,1)$ for some $0 < C < \infty$.*

(b) One of the following conditions hold:

(i) *X is a vector of discrete random variables and $\mathcal{X} \subset \mathbb{R}^{d_X}$ is a finite set.*

(ii) *X is a vector of continuous random variables and $\mathcal{X} \subset \mathbb{R}^{d_X}$ is a bounded, convex set with nonempty interior. For some $c > 0$, $M > 0$, and $\alpha > d_X$,*

$$\mathcal{H} = \{p_{y|x} : \mathcal{X} \to [0,1]^{|\mathcal{Y}|} : p_{y|x}(y|\cdot) \in \mathcal{C}_M^\alpha(\mathcal{X}), y \in \mathcal{Y},\ p_{y|x}(y|x) \geq c > 0, \forall (y,x) \in \mathcal{Y} \times \mathcal{X}\}$$

(iii) *$X = (X_c^\top, X_d^\top)^\top$ consists of subvectors X_c and X_d, where X_c is continuously distributed and X_d is discretely distributed. $\mathcal{X} = \mathcal{X}_c \times \mathcal{X}_d \subset \mathbb{R}^{d_X}$, where $\mathcal{X}_c \subset \mathbb{R}^{d_{X_c}}$ is a bounded convex set with nonempty interior, and $\mathcal{X}_d \subset \mathbb{R}^{d_{X_d}}$ is a finite set. For some $c > 0$, $M > 0$, $\alpha > d_X$, Lipschitz functions $\phi_k, k = 1, \ldots, |\mathcal{Y}|$, and some functions ℓ_c and ℓ_d,*

$$\begin{aligned}\mathcal{H} = \{p_{y|x} : \mathcal{X} \to [0,1]^{|\mathcal{Y}|} : p_{y|x}(y_k|x) = \phi_k\big(\ell_c(y_k|x_c), \ell_d(y_k|x_d)\big),\ \ell_c(y_k|\cdot) \in \mathcal{C}_M^\alpha(\mathcal{X}_c),\\ -M \leq \ell_d(y_k|x_d) \leq M, \forall x_d \in \mathcal{X}_d,\ k = 1, \ldots, d_Y,\ p_{y|x}(y|x) \geq c > 0, \forall (y,x) \in \mathcal{Y} \times \mathcal{X}\}.\end{aligned}$$

REMARK B.1: Assumption B.1(i) is satisfied, for example, when the vector U has a multivariate Normal distribution, provided the correlation among any two of its entries is bounded away from one (in absolute value). In Assumption B.2(b)(iii), we assume that $p_{y|x}$ combines a function of continuous covariates X_c with a function of discrete covariates X_d using a Lipschitz function, which covers many transformations of interest (see, e.g., van der Vaart and Wellner, 1996, p. 192). More general transformations can be allowed for, as far as one may ensure that the metric entropy of $\mathcal{H}$ can be controlled properly.

PROPOSITION B.3: *Suppose Assumptions 1 and B.1 hold for the entry game model in Example 1. Then Assumption 2 also holds.*

PROOF: Recall $m_\theta(x; p_{y|x}) \equiv \mathbb{E}[s_\theta(Y|X; p_{y|x})|X = x] = \sum_{y \in \mathcal{Y}} p_{0,y|x}(y|x) s_\theta(y, x; p_{y|x})$. For $p_{y|x}, p_{0,y|x} \in \mathcal{H}$, we aim to bound $\mathbb{E}[\|m_\theta(X; p_{y|x}) - m_\theta(X; p_{0,y|x})\|]$. The score depends on $p_{y|x}$ only through $\mathbb{I}(x; p_{y|x}) = (\mathbb{I}_1(x; p_{y|x}), \mathbb{I}_2(x; p_{y|x}),\ \mathbb{I}_3(x; p_{y|x}))$. Hence, $\Delta(x; p_{y|x}, p_{0,y|x}) \equiv \|m_\theta(x; p_{y|x}) - m_\theta(x; p_{0,y|x})\| \neq 0$ only if $\mathbb{I}(x; p_{y|x}) \neq \mathbb{I}(x; p_{0,y|x})$. The

TABLE B.I

VALUES OF $\Delta(x;p_{y|x},p_{0,y|x})$ WHEN $\mathbb{I}(x;p_{y|x})\neq\mathbb{I}(x;p_{0,y|x})$

$\mathbb{I}(x;p_{y\|x})$	$\mathbb{I}(x;p_{0,y\|x})$	$\Delta(x;p_{y\|x},p_{0,y\|x})$
(1,0,0) (0,1,0)	(0,1,0) (1,0,0)	$\Big\|p_{0,y\|x}((1,0)\|x)\Big(\frac{\nabla_\theta\eta_2(\theta;x)}{\eta_2(\theta;x)}-\frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\Big)+p_{0,y\|x}((0,1)\|x)\Big(\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)-\eta_2(\theta;x)}-\frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\Big)\Big\|$
(1,0,0) (0,0,1)	(0,0,1) (1,0,0)	$\Big\|p_{0,y\|x}((1,0)\|x)\Big(\frac{\nabla_\theta\eta_3(\theta;x)}{\eta_3(\theta;x)}-\frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\Big)+p_{0,y\|x}((0,1)\|x)\Big(\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_3(\theta;x)}{\eta_1(\theta;x)-\eta_3(\theta;x)}-\frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\Big)\Big\|$
(0,1,0) (0,0,1)	(0,0,1) (0,1,0)	$\Big\|p_{0,y\|x}((1,0)\|x)\Big(\frac{\nabla_\theta\eta_2(\theta;x)}{\eta_2(\theta;x)}-\frac{\nabla_\theta\eta_3(\theta;x)}{\eta_3(\theta;x)}\Big)+p_{0,y\|x}((0,1)\|x)\Big(\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)-\eta_2(\theta;x)}-\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_3(\theta;x)}{\eta_1(\theta;x)-\eta_3(\theta;x)}\Big)\Big\|$

values of $\Delta(x;p_{y|x},p_{0,y|x})$ are given in Table B.I. We consider two subcases (i) X is discrete and $\mathcal{X}$ is finite, and (ii) X contains a continuously distributed variable (the case where all components of X are continuously distributed is treated as a special case of the latter).

(i) Discrete X**:** Let $\mathcal{X}$ be a finite set, $\mathcal{X}_0=\{x\in\mathcal{X}:Z_1(x;p_{0,y|x})\neq 0, Z_2(x;p_{0,y|x})\neq 0\}$, and $c\equiv\min_{x\in\mathcal{X}_0}\min_{j=1,2}|Z_j(x;p_{0,y|x})|$. By Lemma B.4, $\Delta(x;p_{y|x},p_{0,y|x})=0$ for all $x\in\mathcal{X}_0$ and $p_{y|x}$ such that $\|p_{y|x}-p_{0,y|x}\|_{\mathcal{H}}\leq\delta$ for all $\delta\leq c/4$, hence they do not contribute to the L^1-norm of $\Delta(\cdot;p_{y|x},p_{0,y|x})$. Take $x\in\mathcal{X}_0^c$. Suppose, e.g., that $0=Z_1(x;p_{0,y|x})<Z_2(x;p_{0,y|x})$. To have $\Delta(x;p_{y|x},p_{0,y|x})\neq 0$, let $p_{y|x}$ be such that $\|p_{y|x}-p_{0,y|x}\|_{\mathcal{H}}\leq\delta$ and $Z_1(x;p_{y|x})>0$. This yields $\mathbb{I}(x,p_{y|x})=(0,1,0)$ and $\mathbb{I}(x,p_{0,y|x})=(1,0,0)$. Per Table B.I,

$$\Delta(x;p_{y|x},p_{0,y|x})=\Big\|p_{0,y|x}((1,0)|X)\frac{\nabla_\theta\eta_2(\theta;X)}{\eta_2(\theta;X)}+p_{0,y|x}((0,1)|X)\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)-\eta_2(\theta;x)} \\ -[p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X)]\frac{\nabla_\theta\eta_1(\theta;X)}{\eta_1(\theta;X)}\Big\|. \quad \text{(B.12)}$$

Note that $Z_1(x;p_{0,y|x})=0$ is equivalent to

$$p_{0,y|x}((1,0)|X)=[p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X)]\frac{\eta_2(\theta;X)}{\eta_1(\theta;X)}. \quad \text{(B.13)}$$

Since $\frac{p_{0,y|x}((0,1)|X)}{p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X)}=1-\frac{p_{0,y|x}((1,0)|X)}{p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X)}$, we obtain

$$p_{0,y|x}((0,1)|X)=[p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X)]\frac{\eta_1(\theta;X)-\eta_2(\theta;X)}{\eta_1(\theta;X)}. \quad \text{(B.14)}$$

Substituting (B.13)-(B.14) into (B.12) yields $\Delta(x;p_{y|x},p_{0,y|x}) = \Big\|[p_{0,y|x}((1,0)|X) + p_{0,y|x}((0,1)|X)]\Big(\frac{\nabla_\theta\eta_2(\theta;X)}{\eta_1(\theta;X)} + \frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)} - \frac{\nabla_\theta\eta_1(\theta;X)}{\eta_1(\theta;X)}\Big)\Big\| = 0$. A similar argument can be applied to $x \in \mathcal{X}_0^c$ such that $Z_1(x;p_{0,y|x}) < Z_2(x;p_{0,y|x}) = 0$. Finally, consider $x \in \mathcal{X}_0^c$ such that $Z_1(x;p_{0,y|x}) = Z_2(x;p_{0,y|x}) = 0$. This occurs only if $\eta_2(\theta;x) = \eta_3(\theta;x)$. Hence, $Z_1(x;p_{y|x}) = Z_2(x;p_{y|x})$ for any $p_{y|x}$. It then suffices to consider only one of Z_j's. For example let $p_{y|x}$ be such that $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \le \delta$ and $Z_1(x;p_{y|x}) > 0$. Then, the same analysis as above leads to $\Delta(x;p_{y|x},p_{0,y|x}) = 0$. Therefore, for all $p_{y|x}$ such that $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \le \delta$ for a sufficiently small δ, the pathwise derivative is 0, as

$$\Big\|\mathbb{E}\Big[m_\theta(X;p'_{y|x}) - m_\theta(X;p_{0,y|x})\Big]\Big\| \le \mathbb{E}\Big[\|m_\theta(X;p'_{y|x}) - m_\theta(X;p_{0,y|x})\|\Big]$$
$$= \sum_{x\in\mathcal{X}_0} p_{0,x}(x)\Delta(x;p_{y|x},p_{0,y|x}) + \sum_{x\in\mathcal{X}_0^c} p_{0,x}(x)\Delta(x;p_{y|x},p_{0,y|x}) = 0.$$

(ii) X contains a continuously distributed variable:

Let X_d be a subvector of X containing discrete covariates. Recall that $\Delta(x,p_{y|x},p_{0,y|x}) \neq 0$ when $\mathbb{I}(x;p_{y|x}) \neq \mathbb{I}(x;p_{0,y|x})$. This occurs when $\operatorname{sgn}(Z_j(x;p_{y|x})) \neq \operatorname{sgn}(Z_j(x;p_{0,y|x}))$ for some j. By (B.8) and $\sup_{x\in\mathcal{X}}|\eta_j(\theta;x)| \le 1$, $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \le \delta$ implies

$$\sup_{x\in\mathcal{X}} |Z_j(x;p_{y|x}) - Z_j(x;p_{0,y|x})| \le 3\delta,\ j = 1,2$$

Therefore, if $\operatorname{sgn}(Z_j(x;p_{y|x})) \neq \operatorname{sgn}(Z_j(x;p_{0,y|x}))$ and $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \le \delta$, one must have $|Z_j(x;p_{0,y|x})| \le 3\delta$. Hence, the pathwise derivative is again zero because

$$\Big\|E\Big[m_\theta(X;p_{y|x}) - m_\theta(X;p_{0,y|x})\Big]\Big\| \le \mathbb{E}\Big[\|m_\theta(X;p_{y|x}) - m_\theta(X;p_{0,y|x})\|\Big]$$
$$\le \mathbb{E}\Big[\Delta(X;p_{y|x},p_{0,y|x})(1\{-3\delta \le Z_1(X;p_{0,y|x}) \le 3\delta\} + 1\{-3\delta \le Z_2(X;p_{0,y|x}) \le 3\delta\})\Big]$$
$$\overset{(i)}{\le} K\delta\mathbb{E}\Big[\int 1\{-3\delta \le z_1 \le 3\delta\} f_{Z_1|X_d}(z_1)dz_1 + \int 1\{-3\delta \le z_2 \le 3\delta\} f_{Z_2|X_d}(z_2)dz_2\Big] \overset{(ii)}{\le} c\delta^2,$$

where inequality (i) follows by Lemma B.5 and the law of iterated expectations, and inequality (ii) follows by Assumption B.1 (iii) for $0 < c < \infty$ some constant. *Q.E.D.*

LEMMA B.4: *Let $\mathcal{X}$ be a finite set, and let $\mathcal{X}_0 = \{x \in \mathcal{X} : Z_1(x; p_{0,y|x}) \neq 0, Z_2(x; p_{0,y|x}) \neq 0\}$. Let $c \equiv \min_{x\in\mathcal{X}_0} \min_{j=1,2} |Z_j(x; p_{0,y|x})|$. Then, $\Delta(x; p_{y|x}, p_{0,y|x}) = 0$ for any $x \in \mathcal{X}_0$ and $p_{y|x}$ such that $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \leq c/4$.*

PROOF: Take $x \in \mathcal{X}_0$. Suppose $Z_1(x; p_{0,y|x}) \geq c > 0$ so that $\mathbb{I}(x; p_{0,y|x}) = (0, 1, 0)$. Let $p_{y|x}$ satisfy $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \leq c/4$. Then, by (B.8) and $|\eta_j(x; \theta)| \leq 1$, and the triangle inequality, $Z_1(x; p_{y|x}) \geq Z_1(x; p_{0,y|x}) - \frac{3}{4}c \geq \frac{1}{4}c > 0$, implying $\mathbb{I}(x; p_{y|x}) = (0, 1, 0)$. From Table B.I, $\Delta(x; p_{y|x}, p_{0,y|x}) = 0$. Other cases can be analyzed similarly. *Q.E.D.*

LEMMA B.5: *Suppose Assumptions 1 and B.1 hold for the entry game model in Example 1. For $\delta > 0$, let $p_{y|x}$ be such that $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \leq \delta$. Then, there exists $0 < K < \infty$ such that for all $x \in \mathcal{X}$, $\Delta(x; p_{y|x}, p_{0,y|x}) \leq K\delta$.*

PROOF: From Table B.I, $\Delta(x; p_{y|x}, p_{0,y|x}) = 0$ when $\mathbb{I}(x; p_{y|x}) = \mathbb{I}(x; p_{0,y|x})$. Therefore, we focus on cases with $\mathbb{I}(x; p_{y|x}) \neq \mathbb{I}(x; p_{0,y|x})$ below. Consider the case where $\mathbb{I}(x; p_{y|x}) = (0, 1, 0)$ and $\mathbb{I}(x; p_{0,y|x}) = (1, 0, 0)$. By (B.9)-(B.11), this occurs when

$$Z_1(x; p_{0,y|x}) \leq 0, \ Z_1(x; p_{y|x}) > 0. \tag{B.15}$$

Furthermore, by (B.8) and $\sup_{x\in\mathcal{X}} |\eta_j(\theta; x)| \leq 1$, $\|p_{y|x} - p_{0,y|x}\|_{\mathcal{H}} \leq \delta$ implies

$$\sup_{x\in\mathcal{X}} |Z_1(x; p_{y|x}) - Z_1(x; p_{0,y|x})| \leq 3\delta. \tag{B.16}$$

Combining (B.15)-(B.16) yields $-3\delta \leq Z_1(x; p_{0,y|x}) \leq 0$. By (B.8) and $\eta_j(\theta; x) \geq c$,

$$\begin{aligned}(p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\tfrac{\eta_2(\theta;x)}{\eta_1(\theta;x)} - \tfrac{3}{c}\delta \\ \leq p_{0,y|x}((1,0)|x) \leq (p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\tfrac{\eta_2(\theta;x)}{\eta_1(\theta;x)}\end{aligned} \tag{B.17}$$

Using Assumption B.1 (ii), this may also be written as

$$\tfrac{\eta_2(\theta;x)}{\eta_1(\theta;x)} - \tfrac{3}{c(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))}\delta \leq \tfrac{p_{0,y|x}((1,0)|x)}{(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))} \leq \tfrac{\eta_2(\theta;x)}{\eta_1(\theta;x)}.$$

Since $\frac{p_{0,y|x}((0,1)|x)}{(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))} = 1 - \frac{p_{0,y|x}((1,0)|x)}{(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))}$, we obtain

$$1 - \frac{\eta_2(\theta;x)}{\eta_1(\theta;x)} \leq \frac{p_{0,y|x}((0,1)|x)}{(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))} \leq 1 - \frac{\eta_2(\theta;x)}{\eta_1(\theta;x)} + \frac{3}{c(p_{0,y|x}((1,0)|X)+p_{0,y|x}((0,1)|X))}\delta.$$

This may in turn be written as

$$\begin{aligned}(p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\frac{\eta_1(\theta;x)-\eta_2(\theta;x)}{\eta_1(\theta;x)}&\\ \leq p_{0,y|x}((0,1)|x) \leq (p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\frac{\eta_1(\theta;x)-\eta_2(\theta;x)}{\eta_1(\theta;x)} + \frac{3}{c}\delta.&\end{aligned} \tag{B.18}$$

By (B.17) and (B.18), let us write

$$p_{0,y|x}((1,0)|x) = (p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\frac{\eta_2(\theta;x)}{\eta_1(\theta;x)} + r_{(1,0)}(x) \tag{B.19}$$

$$p_{0,y|x}((0,1)|x) = (p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x))\frac{\eta_1(\theta;x)-\eta_2(\theta;x)}{\eta_1(\theta;x)} + r_{(0,1)}(x), \tag{B.20}$$

where $r_{(1,0)}(x) \in [-3\delta/c, 0]$ and $r_{(0,1)}(x) \in [0, 3\delta/c]$ for all $x \in \mathcal{X}$. From Table B.I, the value of $\Delta(x; p_{y|x}, p_{0,y|x})$ when $\mathbb{I}(x; p_{y|x}) = (0,1,0)$ and $\mathbb{I}(x; p_{0,y|x}) = (1,0,0)$ is

$$\begin{aligned}\Delta(x; p_{y|x}, p_{0,y|x}) = \Bigg\| p_{0,y|x}((1,0)|x)\frac{\nabla_\theta\eta_2(\theta;x)}{\eta_2(\theta;x)} + p_{0,y|x}((0,1)|x)\frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)-\eta_2(\theta;x)}&\\ - [p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x)]\frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\Bigg\|.&\end{aligned} \tag{B.21}$$

By (B.19)-(B.20), the terms inside the norm in (B.21) can therefore be written as

$$\begin{aligned}[p_{0,y|x}((1,0)|x) + p_{0,y|x}((0,1)|x)]\left(\frac{\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)} + \frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)} - \frac{\nabla_\theta\eta_1(\theta;x)}{\eta_1(\theta;x)}\right)&\\ + \frac{\nabla_\theta\eta_2(\theta;x)}{\eta_2(\theta;x)}r_{(1,0)}(x) + \frac{\nabla_\theta\eta_1(\theta;x)-\nabla_\theta\eta_2(\theta;x)}{\eta_1(\theta;x)-\eta_2(\theta;x)}r_{(0,1)}(x).&\end{aligned}$$

By the triangle inequality, $\eta_j(\theta;x) \geq c$, and Assumption B.1 (i), we obtain

$$\Delta(x; p_{y|x}, p_{0,y|x}) \leq \|\nabla_\theta\eta_2(\theta;x)\|\left|\frac{r_{(1,0)}(x)}{\eta_2(\theta;x)}\right| + (\|\nabla_\theta\eta_1(\theta;x)\| + \|\nabla_\theta\eta_2(\theta;x)\|)\left|\frac{r_{(0,1)}(x)}{\eta_1(\theta;x)-\eta_2(\theta;x)}\right|$$

$$\leq \tfrac{3C}{c^2}\delta + \tfrac{6C}{c^2}\delta,$$

which establishes the claim of the lemma for $\mathbb{I}(x;p_{y|x}) = (0,1,0)$ and $\mathbb{I}(x;p_{0,y|x}) = (1,0,0)$. The other cases can be analyzed similarly. *Q.E.D.*

We next establish stochastic equicontinuity for the empirical process

$$\mathbb{G}_n(p) = \tfrac{1}{\sqrt{n}}\sum_{i=1}^n \left(s_\theta(Y_i|X_i;p) - \mathbb{E}[s_\theta(Y_i|X_i;p)]\right), \ \ p \in \mathcal{H}$$

In the definition of $\mathbb{G}_n$, the structural parameter θ is fixed. Hence, the function class $\mathcal{F} = \{f(y,x) : f(y,x) = s_\theta(y|x;p), p \in \mathcal{H}\}$ is defined by mixing and matching p with fixed functions such as $\eta_j(x,\theta)$. For example, we may write $s_\theta((1,0)|x;p)$ as

$$s_\theta((1,0)|x;p_{0,y|x}) = \sum_{j=1}^3 \tfrac{\nabla_\theta \eta_j(\theta;x)}{\eta_j(\theta;x)} \mathbb{I}_j(x;p_{y|x}),$$

where $\mathbb{I}_j(x;p_{y|x}), j = 1,2,3$, are defined in (B.9)-(B.11).

PROPOSITION B.4: *Suppose Assumptions 1, B.1, and B.2 hold for the entry game model in Example 1. Then Assumption 3 (iv) also holds.*

PROOF: Let

$$\mathcal{F}_{(0,0)} = \Big\{ f : f(w;p) = \tfrac{\nabla_\theta F_\theta(S_{\{(0,0)\}|x;\theta})}{F_\theta(S_{\{(0,0)\}|x;\theta})} \Big\},$$

$$\mathcal{F}_{(0,1)} = \Big\{ f : f(w;p) = \tfrac{e_l' \nabla_\theta \eta_1(\theta;x)}{\eta_1(\theta;x)} \mathbb{I}_1(x;p) + \sum_{j=2}^3 \tfrac{e_l'(\nabla_\theta \eta_1(\theta;x) - \nabla_\theta \eta_j(\theta;x))}{\eta_j(\theta;x)} \mathbb{I}_j(x;p), l = 1,\dots,d_\theta, \ p \in \mathcal{H} \Big\},$$

$$\mathcal{F}_{(1,0)} = \Big\{ f : f(w;p) = \sum_{j=1}^3 \tfrac{e_l' \nabla_\theta \eta_j(\theta;x)}{\eta_j(\theta;x)} \mathbb{I}_j(x;p), \ l = 1,\dots,d_\theta, \ p \in \mathcal{H} \Big\},$$

$$\mathcal{F}_{(1,1)} = \Big\{ f : f(w;p) = \tfrac{\nabla_\theta F_\theta(S_{\{(1,1)\}|x;\theta})}{F_\theta(S_{\{(1,1)\}|x;\theta})} \Big\}$$

where for each l, e_l denotes the l-th standard basis vector in $\mathbb{R}^{d_\theta}$.

The score satisfies $s_\theta(\cdot;p_{y|x}) \in \mathcal{F} \equiv \sum_{\bar{y}\in\mathcal{Y}} \mathcal{F}_{\bar{y}} \cdot 1\{y=\bar{y}\}$. In view of Theorem 2.10.6 (and Examples 2.10.7 and 2.10.10) in van der Vaart and Wellner (1996), to verify Assumption 3 (iv) it suffices to show that $\mathcal{F}_{\bar{y}}$ is P-Donsker for each $\bar{y}$. The P-Donskerness of $\mathcal{F}_{\bar{y}}$ for $\bar{y}=(0,0),(1,1)$ follows from each set being a singleton and Assumption B.2 (a). Below, we show $\mathcal{F}_{(1,0)}$ is P-Donsker. The analysis for $\mathcal{F}_{(0,1)}$ is similar and is therefore omitted.

Discrete X: First, suppose that X is a vector of discrete random variables and Assumption B.2(b)(i) holds. We show that $\mathcal{F}_{(1,0)}$ is a Vapnik-Chervonenkis (VC) class, which satisfies Pollard's uniform entropy condition. As $\mathcal{Y}\times\mathcal{X}$ is finite, $\mathcal{H}$ is finite-dimensional. By van der Vaart and Wellner (1996, Lemma 2.6.15), this class has VC-index $V(\mathcal{H}) \le d_Y \times d_X + 2$, and hence $\mathcal{H}$ is a VC-class. The finite set of functions $\mathcal{E} = \{\eta_j(\theta,\cdot), \frac{\partial}{\partial\theta_k}\eta_j(\theta,\cdot), j = 1,\dots,3, k=1,\dots,d_\theta\}$ is also a VC-class. As $\mathcal{F}_{(1,0)}$ collects functions that can be expressed as combinations of functions from $\mathcal{H}$ and $\mathcal{E}$ by multiplication, addition, division, and composition with an indicator function $1\{\cdot>0\}$, $\mathcal{F}_{(1,0)}$ is a VC-class (van der Vaart and Wellner, 1996, Lemma 2.6.18). Assumptions 1 and B.1(i) ensure that there is an envelope (constant) function $F = 3C/c$ such that $|f| \le F$ for all $f \in \mathcal{F}_{(1,0)}$. By Theorem 2.5.2 in van der Vaart and Wellner (1996), $\mathcal{F}_{(1,0)}$ is a P-Donsker class.

Continuous X: Next, suppose that X is a vector of continuous random variables and Assumption B.2(b)(ii) holds. We show $\mathcal{F}_{(1,0)}$ is P-Donsker by verifying the conditions in Chen et al. (2003, Theorem 3). We first show the L^2-Hölder continuity of $f \in \mathcal{F}_{(1,0)}$ in p. In what follows, let $U_{l,j} = e_l'\nabla_\theta\eta_j(\theta;X)/\eta_j(\theta;X)$, $l=1,\dots,d$, $j=1,\dots,3$ and $u_{l,j} = e_l'\nabla_\theta\eta_j(\theta;x)/\eta_j(\theta;x)$. By the triangle inequality,

$$\sup_{\|p-p'\|_{\mathcal{H}}\le\delta} |f(w;p') - f(w;p)|^2 \le \sup_{\|p-p'\|_{\mathcal{H}}\le\delta} \sum_{j=1}^{3} u_{l,j}^2 |\mathbb{I}_j(x;p') - \mathbb{I}_j(x;p)|, \tag{B.22}$$

Below, we focus on $u_{l,3}^2|\mathbb{I}_3(x;p') - \mathbb{I}_3(x;p)|$, one of the terms in the sum on the right hand side of (B.22). The two other terms can be analyzed similarly. For δ sufficiently small,

$$\mathbb{E}\Big[\sup_{\|p-p'\|_{\mathcal{H}}\le\delta} U_{l,3}^2|\mathbb{I}_3(X;p') - \mathbb{I}_3(X;p)|\Big]$$

$$= \mathbb{E}\Big[\sup_{\|p-p'\|_{\mathcal{H}}\le\delta} U_{l,3}^2 |1\{Z_2(X;p'_{y|x})<0\} - 1\{Z_2(X;p_{y|x})<0\}|\Big].$$

By (B.8) and $\sup_{x\in\mathcal{X}}|\eta_j(\theta;x)|\le 1$, whenever $\|p'_{y|x}-p_{y|x}\|_{\mathcal{H}}\le\delta$, we have

$$\sup_{x\in\mathcal{X}}|Z_j(x;p'_{y|x})-Z_j(x;p_{y|x})|\le 3\delta,\ j=1,2. \tag{B.23}$$

We next use the argument in Chen et al. (2003, p. 1600). Combining one side of (B.23), with the addition of a non-negative constant, we have $Z_2(x;p_{y|x})-3\delta\le Z_2(x;p'_{y|x})\le Z_2(x;p'_{y|x})+3\delta$, and hence

$$1\{Z_2(x;p_{y|x})-3\delta<0\}\ge 1\{Z_2(x;p'_{y|x})<0\}\ge 1\{Z_2(x;p'_{y|x})+3\delta<0\}. \tag{B.24}$$

Similarly, $Z_2(x;p'_{y|x})-3\delta\le Z_2(x;p_{y|x})\le Z_2(x;p_{y|x})+3\delta$ implies

$$1\{Z_2(x;p'_{y|x})-3\delta<0\}\ge 1\{Z_2(x;p_{y|x})<0\}\ge 1\{Z_2(x;p_{y|x})+3\delta<0\}. \tag{B.25}$$

Combining (B.24)-(B.25), for any p',p with $\|p'-p\|_{\mathcal{H}}\le\delta$,

$$\begin{aligned}|1\{Z_2(x;p'_{y|x})<0\}-1\{Z_2(x;p_{y|x})<0\}| &\le 1\{Z_2(x;p_{y|x})-3\delta<0\}-1\{Z_2(x;p_{y|x})+3\delta<0\}\\ &\le 1\{-3\delta<Z_2(x;p_{y|x})<3\delta\}\end{aligned}$$

where without loss of generality we assumed that $1\{Z_2(x;p_{y|x})-3\delta<0\}-1\{Z_2(x;p_{y|x})+3\delta<0\}>1\{Z_2(x;p'_{y|x})-3\delta<0\}-1\{Z_2(x;p'_{y|x})+3\delta<0\}$. By the argument above, the law of iterated expectations, and Assumptions B.1(i), B.2(b)(ii), and B.2(a),

$$\begin{aligned}&\mathbb{E}\Big[\sup_{\|p-p'\|_{\mathcal{H}}\le\delta} U_{l,3}^2|1\{Z_2(X;p'_{y|x})<0\}-1\{Z_2(X;p_{y|x})<0\}|\Big]\\ &\qquad\le \mathbb{E}\Big[U_{l,3}^2 1\{-3\delta<Z_2(X;p_{y|x})<3\delta\}\Big]\\ &\qquad\le \frac{C^2}{c^2}\int 1\{-3\delta<z_2<3\delta\} f_{Z_2}(z_2)dz_2\le K\delta,\end{aligned} \tag{B.26}$$

for some constant $K>0$, where the last inequality follows from Assumption B.1 (iii). Applying a similar argument to the other two terms in (B.22), one can obtain

$$\mathbb{E}\Big[\sup_{\|p-p'\|_{\mathcal{H}}\le\delta}|f(W;p')-f(W;p)|^2\Big]^{1/2}\le K'\delta^{1/2}, \tag{B.27}$$

for some $K'>0$. Hence f is L^2-Hölder continuous in p with Hölder exponent $1/2$.

Recall that $\mathcal{X}$ is a bounded convex subset of $\mathbb{R}^{d_X}$ with nonempty interior. By Theorem 2.7.1 in van der Vaart and Wellner (1996), $\ln N(\epsilon^2,\mathcal{C}^\alpha_M(\mathcal{X}),\|\cdot\|_\infty)\le K\left(\frac{1}{\epsilon}\right)^{2d_X/\alpha}$ for some $K>0$. Note that $\mathcal{H}\subset(\mathcal{C}^\alpha_M(\mathcal{X}))^{\mathcal{Y}}$ and $|\mathcal{Y}|=4$. For each $y\in\mathcal{Y}$, let $\{p_1(y|\cdot),\dots,p_k(y|\cdot)\}$ be an ϵ^2-cover for $\mathcal{C}^\alpha_M(\mathcal{X})$ with respect to the sup norm. Then, $\{(p_{j_1}((0,0)|\cdot),p_{j_2}((0,1)|\cdot),p_{j_3}((1,0)|\cdot),$ $p_{j_4}((1,1)|\cdot),j_l\in\{1,\dots,k\},l=1,\dots,4\}$ forms an ϵ^2-cover for $(\mathcal{C}^\alpha_M(\mathcal{X}))^{\mathcal{Y}}$ with respect to the maximum of the sup norms. Hence,

$$N(\epsilon^2,\mathcal{C}^\alpha_M(\mathcal{X})^{\mathcal{Y}},\|\cdot\|_\infty)\le e^{4K\left(\frac{1}{\epsilon}\right)^{2d_X/\alpha}}, \tag{B.28}$$

which in turn implies $\ln N(\epsilon^2,\mathcal{H},\|\cdot\|_\infty)\le 4K\left(\frac{1}{\epsilon}\right)^{2d_X/\alpha}$. Since $\alpha>d_X$, we have

$$\int_0^\infty\sqrt{\ln N(\epsilon^2,\mathcal{H},\|\cdot\|_{\mathcal{H}})}d\epsilon<\infty. \tag{B.29}$$

We can now apply Theorem 3 in Chen et al. (2003), which ensures that $\mathcal{F}_{(1,0)}$ is P-Donsker.

Mixed X: Finally, suppose that X contains both continuous and discrete variables and Assumption B.2(b)(iii) holds. Again, we use Theorem 3 in Chen et al. (2003). We can argue as in the previous case, but (B.26) is modified as follows:

$$\begin{aligned}\mathbb{E}\Big[\sup_{\|p-p'\|_{\mathcal{H}}\le\delta}&U^2_{l,3}|1\{Z_2(X;p'_{y|x})<0\}-1\{Z_2(X;p_{y|x})<0\}|\Big]\\&\le\mathbb{E}\Big[U^2_{l,3}1\{-3\delta<Z_2(X;p_{y|x})<3\delta\}\Big]\\&\le\frac{C^2}{c^2}\mathbb{E}\left[\int1\{-3\delta<z_2<3\delta\}f_{X_2|X_d}(z_2)dz_2\right]\le K\delta,\end{aligned}\tag{B.30}$$

for some constant $K > 0$, where the last inequality follows from Assumption B.1(iii). Therefore, (B.27) holds.

Next we show (B.29). Recall that $N(\epsilon^2, \mathcal{C}_M^\alpha(\mathcal{X}), \|\cdot\|_\infty) \leq e^{K\left(\frac{1}{\epsilon}\right)^{2d_X/\alpha}}$ for some $K > 0$. Furthermore, $x_d \mapsto \ell_d(y_k|x_d)$ belongs to a finite-dmensional space $[-M, M]^{\mathcal{X}_d}$ with covering number $N(\epsilon^2, [-M, M]^{\mathcal{X}_d}, \|\cdot\|_\infty) \leq \left(\frac{\sqrt{2M}}{\epsilon}\right)^{2\dim(\mathcal{X}_d)}$. For each l, let $p_{c,1}(y_l|\cdot), \cdots, p_{c,N_1}(y_l|\cdot)$ be an ϵ^2-cover of $\mathcal{C}_M^\alpha(\mathcal{X}_c)$. Similarly, let $p_{d,1}(y_l|\cdot), \ldots, p_{d,N_2}(y_l|\cdot)$ be an ϵ^2-cover of $[-M, M]^{\mathcal{X}_d}$. Then, for any $p_{y|x} \in \mathcal{H}$ and $l \in \{1, \ldots, 4\}$, there exist $k_1 \in \{1, \ldots, N_1\}$, $k_2 \in \{1, \ldots, N_2\}$, and $(\ell_c(y_k|\cdot), \ell_d(y_k|\cdot)) \in \mathcal{C}_M^\alpha(\mathcal{X}_c) \times [-M, M]^{\mathcal{X}_d}$ such that

$$
\begin{aligned}
&\sup_{x=(x_c', x_d')' \in \mathcal{X}_c \times \mathcal{X}_d} \left| p_{y_l|x}(y|x) - \phi_k(p_{c,k_1}(y_l|x_c), p_{d,k_1}(y_l|x_d)) \right| \\
&\quad = \sup_{x=(x_c', x_d')' \in \mathcal{X}_c \times \mathcal{X}_d} \left| \phi_k(p_c(y_l|x_c), p_d(y_l|x_d)) - \phi_k(p_{c,k_1}(y_l|x_c), p_{d,k_2}(y_l|x_d)) \right| \\
&\qquad\qquad \leq C \max\{\|p_c(y_l|\cdot) - p_{c,k_1}(y_l|\cdot)\|_\infty, \|p_d(y_l|\cdot) - p_{d,k_2}(y_l|\cdot)\|_\infty\} \leq C\epsilon^2,
\end{aligned}
$$

for some $0 < C < \infty$ due to the Lipschitz continuity of ϕ_k. Therefore $\{(p_{c,k_1}(y_l|\cdot), p_{d,k_2}(y_l|\cdot))_{l=1}^4$, $k_1 \in \{1, \ldots, N_1\}, k_2 \in \{1, \ldots, N_2\}, l \in \{1, \ldots, 4\}\}$ is an $C\epsilon^2$-cover of $\mathcal{H}$. Hence,

$$
N(\epsilon^2, \mathcal{H}, \|\cdot\|_\infty) \leq \Big(N(\epsilon^2/C, \mathcal{C}_M^\alpha(\mathcal{X}), \|\cdot\|_\infty) \times N(\epsilon^2/C, [-M, M]^{\mathcal{X}_d}, \|\cdot\|_\infty) \Big)^4,
$$

which in turn implies, for some $K' > 0$ for all ϵ small enough,

$$
\ln N(\epsilon^2, \mathcal{H}, \|\cdot\|_\infty) \leq 4K \left(\frac{\sqrt{C}}{\epsilon}\right)^{2d_X/\alpha} + 8\dim(\mathcal{X}_d) \ln\left(\frac{\sqrt{2M}}{\epsilon}\right) \leq K' \left(\frac{\sqrt{C}}{\epsilon}\right)^{2d_X/\alpha}.
$$

Again, by $\alpha > d$, we obtain (B.29). This completes the proof of the proposition. *Q.E.D.*

We conclude this section by arguing that provided $\boldsymbol{X}$ has at least one component with continuous distribution, under Assumptions 3 (ii), B.1 (iii), and B.2 (a), the consistency of the covariance matrix estimator $\hat{\Sigma}_{n,\theta^*}$ required in (3.35) holds. This follows from (B.27), arguing as in Powell et al. (1989, Theorem 3.4), leveraging Assumption 3 (ii) and that for $\bar{y} = (0,0), (1,1)$ the score does not depend on $p_{n,y|x}$ together with Assumption B.2 (a).

B.3. *Uniform Convergence for Series Estimators of* p_0

In this section we provide sufficient conditions under which it is possible to verify Assumption 3(ii)' in Theorem 3.4 through an application of results in Chen and Christensen (2015). A formal verification of Assumption 3(ii)' is available from the authors upon request. For simplicity, we focus on the setting where all components of X are continuous. We let $b^K(x) = (b_{K1}(x), \ldots, b_{KK}(x))' \in \mathbb{R}^K$ be a collection of K basis functions, and let $B = (b^K(X_1), \ldots, b^K(X_n))' \in \mathbb{R}^{n\times K}$. We let $\mathbf{1}_y = (1\{Y_1 = y\}, \ldots, 1\{Y_n = y\})'$. We approximate each $p_0(y|\cdot)$ by the series estimator:

$$\hat{p}_n(y|x) = b^K(x)'(B'B)^- B'\mathbf{1}_y.$$

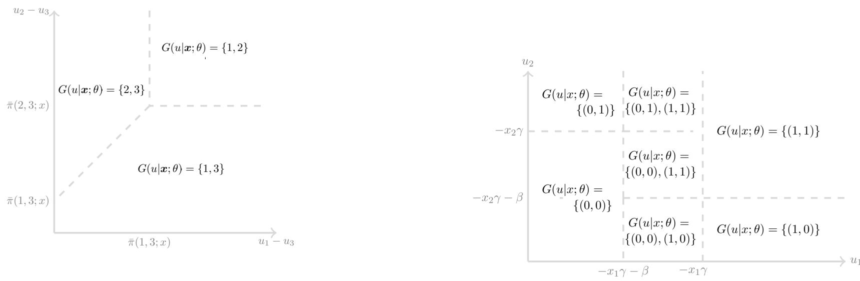
We let $\mathrm{BSpl}(K, [0,1]^{d_X}, \gamma)$ denote a B-spline sieve of degree γ and dimension K on the domain $[0,1]^{d_X}$, and $\mathrm{Wav}(K, [0,1]^{d_X}, \gamma)$ denote a Wavelet sieve basis of regularity γ and dimension K on the domain $[0,1]^{d_X}$. The construction of these sieve spaces is discussed in Chen and Christensen (2015, Section 6). We maintain the following restrictions:

ASSUMPTION B.3: ***(1)*** *$\mathcal{X} = [0,1]^{d_X}$ and $p_{0,X}$ is uniformly bounded away from zero on $\mathcal{X}$.* ***(2)*** *Assumption B.2-(ii) holds and $\|p_{y|x}\|_{\mathcal{H}} = \sup_{y\in\mathcal{Y}}\sup_{x\in\mathcal{X}}|p_{y|x}(y|x)|$.* ***(3)*** *The sieve space is either BSpl$(K, [0,1]^{d_X}, \gamma)$ or Wav$(K, [0,1]^{d_X}, \gamma)$ with $\gamma > \max\{\alpha, 1\}$.* ***(4)*** *$\lambda_{K,n} \equiv \sup_{P\in\mathcal{P}}[\lambda_{min}(\mathbb{E}_P[b_w^K(X_i)b_w^K(X_i)'])]^{-1/2} \lesssim 1$.*

Assumption B.3-(4) essentially assumes a uniform lower bound on the minimum eigenvalue of $\mathbb{E}_P[b_w^K(X_i)b_w^K(X_i)']$. Under Assumption B.3, one can leverage Theorems 2.1 and 3.4 and Lemma 2.3 in Chen and Christensen (2015) to show that if $K \asymp (n/\ln n)^{d_X/(2\alpha+d_X)}$, then $\|\hat{p}_n - p_{0,y|x}\|_{\mathcal{H}} = O_{\mathcal{P}}\big((n/\ln n)^{-\frac{\alpha}{2\alpha+d_X}}\big)$. In particular, one can attain the desired rate $o_{\mathcal{P}}(n^{-1/4})$ if $\alpha > d_X/2$.

APPENDIX C: ADDITIONAL EXAMPLES

Example C.1 (Discrete choice with unobserved heterogeneity in choice sets)**.** Consider a discrete choice model, with a finite universe of alternatives $\mathcal{J} = \{1, \ldots, J\}$. Let each alternative be characterized by a vector of covariates X_j, which might vary across decision



(a) Heterogeneous choice sets (b) Panel dynamic discrete choice

FIGURE C.1.—Stylized depictions of $G(\cdot|x;\theta)$ in Example C.1 (Panel (a), with $\pi(X_j, U;\theta) = \pi(X_j;\theta) + U_j$, $\mathcal{J} = \{1,2,3\}$, $\kappa = 2$, and $\bar{\pi}(j,k;x) \equiv \pi(x_k;\theta) - \pi(x_j;\theta)$) and Example C.2 (Panel (b), with $\beta \geq 0$).

makers, and let $X = [X_j, j \in \mathcal{J}]$. Let $U \in \mathbb{R}^{d_U}$ denote a vector of decision maker and/or alternative specific unobservable attributes. As in the model proposed by Barseghyan et al. (2021), the decision maker draws a *choice set* $C \subseteq \mathcal{J}$ according to an unknown distribution, with $\mathbf{P}(|C| \geq \kappa) = 1$ for some known $\kappa \geq 2$, and chooses the alternative $Y \in C$ that maximizes over alternatives $j \in \mathcal{J}$ the utility $\pi(X_j, U;\theta)$: $Y \in \arg\max_{j \in C} \pi(X_j, U;\theta)$. The researcher observes (Y, X), but not C, and wishes to learn features of θ and the distribution of U. For given $\theta \in \Theta$ and $x \in \mathcal{X}$, Barseghyan et al. (2021, Lemma A.1) show that the set of model implied optimal choices is a measurable correspondence given by the $J - \kappa + 1$ best alternatives in $\mathcal{J}$, so that $G(U|x;\theta) = \cup_{K \subseteq \mathcal{J}:|K|=\kappa} \left\{\arg\max_{j \in K} \pi(x_j, U;\theta)\right\}$.

We depict it in Panel (a) of Figure C.1, for $|\mathcal{J}| = 3$ and $\pi(x_j, U;\theta) = \pi(X_j;\theta) + U_j$ with $U = (U_j, j \in \mathcal{J})$, as a function of $(u_1 - u_3, u_2 - u_3)$. In this example, if U has full support on $\mathbb{R}^J$, Assumption 1(b) is immediately satisfied with $\mathcal{A}_G = \{K \subset \mathcal{Y} : |K| \geq J - \kappa + 1\}$ because each set of alternatives in $\mathcal{J}$ of size $J - \kappa + 1$ can realize as the $J - \kappa + 1$ best. Assumption 1(c)-(d) can be verified similarly to how they are verified for Example 1.

It is also instructive to think about whether the introduction of a selection mechanism can allow for application of the method proposed in Chen et al. (2018) to this example.[2] Let $\mathbf{P}(Y_i = j|X_i;\theta, R)$ denote the model-implied conditional probability that alternative $j \in \mathcal{J}$

[2]In the case of the entry game in Example 1, the selection mechanism in (3.3) can be integrated out against the distribution of U to obtain a function that plays the role of the nuisance parameter in Chen et al. (2018).

is chosen given X_i and (θ, R), where $R(\cdot; X_i, U_i)$ denotes the conditional probability mass function of C_i given (X_i, U_i). For all $j \in \mathcal{J}$,

$$\mathbf{P}(Y_i = j | X_i; \theta, R) = \int \sum_{K \subseteq \mathcal{J}} \mathbf{1}\left(\arg\max_{k \in K} (\pi(X_k; \theta) + u_k) = j\right) R(K; X_i, u) dF_\theta.$$

To be able to apply Chen et al.'s (2018) method, one needs to further restrict the model and assume that R does not depend on U, in which case $R(\cdot; X_i)$ can come out of the integral. Doing so, however, severely restricts the class of models to which the procedure is applicable, since it requires the distribution of choice sets to be independent of preferences. Important examples of choice set formation mechanisms that violate this requirement include sequential search, rational inattention, and elimination by aspects (when the aspect with respect to which elimination occurs is the unobserved characteristic U_j). □

Example C.1 (Continued – Geometric Properties of $\Theta^*(p_0)$). We now specialize Example C.1 to an instance where the mapping $p \mapsto \Theta^*(p)$ is continuous, as discussed in Section 3.4, but $\Theta^*(p_0)$ shrinks to a singleton as the amount of misspecification increases. Let $\mathcal{J} = \{1, 2, 3\}$ and $X_j = [c_j \quad b_j Z^*]$, $Z^* \in \{z_H, z_L\}$ and $z_L < z_H$, with $c_1 < c_2 < c_3$ and $b_1 > b_2 > b_3$ known constants (so that X_j varies stochastically across decision makers, but non-stochastically across alternatives). Let $\pi(X_j, U; \theta) = -((1-U)b_j Z^* + U(c_j + b_j Z^*))$ and $F_\theta(u) = u^\theta$. This example is inspired by the study in Barseghyan et al. (2021, Figure 1-a, Section 6.1.1) of decision making under risk,[3] but for analytic tractability we use a utility function linear in Z^*. It follows that, given $Z^* = z$ alternative j is chosen if and only if $U > \frac{b_1 - b_3}{c_3 - c_1} z$; henceforth, let $A \equiv \frac{b_1 - b_3}{c_3 - c_1} > 0$, so that $P_0(G = \{1, 2\} | Z^* = z) = 1 - (Az)^\theta \equiv \eta(\theta; z)$. We assume that $(b_j, c_j), j \in \{1, 2, 3\}$ are such that $Az \in (0, 1)$ for both $z = z_L$ and $z = z_H$, and that $P_0(C = \{1, 3\} | Z^* = z, U) = 1$ if $U > Az$, and $P_0(C = \{1, 3\} | Z^* = z, U) = 0.5$ if $U \leq Az$. This implies that the true data generating process satisfies $p_{0,y|z}(1|z) = 1 - (Az)^\theta$, $p_{0,y|z}(2|z) = p_{0,y|z}(3|z) = \frac{1}{2}(Az)^\theta$. We generate misspecification similarly to what we did in Example 1 on p.17: the researcher observes a misclassified covariate Z with $P_0(Z^* = z_H | Z = z) = \kappa(z, \gamma) \equiv (1 - \gamma)\mathbf{1}(z = z_H) + \gamma \mathbf{1}(z = z_L)$.

[3]Here, c_j represents the deductible associated with insurance product j and $b_j Z$ the product's price.

We then have that for $y = 2,3$, $p_{\gamma,y|z}(y|z) = \frac{1}{2}[(1-\gamma)(Az_H)^\theta + \gamma(Az_L)^\theta]$ if $z = z_H$ and $p_{\gamma,y|z}(y|z) = \frac{1}{2}[\gamma(Az_H)^\theta + (1-\gamma)(Az_L)^\theta]$ if $z = z_L$; and $p_{\gamma,y|z}(1|z) = 1 - p_{\gamma,2|x}(y|z) - p_{\gamma,3|x}(y|z)$. If $\gamma = 0$, the model is correctly specified and

$$\Theta_I(p_0) = \left[\max_{z\in\{z_L,z_H\}} \frac{\ln(1-p_{0,y|z}(1|z))}{\ln(Az)}, \min_{z\in\{z_L,z_H\}} \frac{\ln(1-p_{0,y|z}(1|z)-p_{0,y|z}(2|z))}{\ln(Az)}\right]. \qquad \text{(C.1)}$$

The bound in (C.1) has the familiar form of an intersection of intervals. Intuitively, for $\gamma > 0$ this bound remains non-empty provided the distributions $p_{\gamma,y|z}(\cdot|z_L)$ and $p_{\gamma,y|z}(\cdot|z_H)$ are sufficiently compatible with each other, and in that case $\Theta^*(p_\gamma) = \Theta_I(p_\gamma)$ because the misspecification is non-detectable: it is possible that p_γ belongs to the model $\mathfrak{Q}$ as in Definition 3.1. But for γ sufficiently large, the distributions $p_{\gamma,y|z}(\cdot|z_L)$ and $p_{\gamma,y|z}(\cdot|z_H)$ become incompatible, $p_\gamma \notin \mathfrak{Q}$, $\Theta_I(p_\gamma) = \emptyset$, and $\Theta^*(p_\gamma)$ becomes a singleton. In what follows we more formally explain these facts.

In this example, $\mathfrak{q}_{\theta,z} = \{q \in \Delta : q(1) \le \eta(\theta;z),\ q_3 \le 1 - (\theta;z)\}$. Following similar steps to the ones we used for Example 1, one can show that in this example, $q^*_{\theta,y|z}(y|z) = p_{\gamma,y|z}(y|z)$ for $y = 1,2,3$ if $\theta \in \Theta_1(z;p_{\gamma,y|z}) \equiv \{\theta : p_{\gamma,y|z}(1|z) \le \eta(\theta;z) \le p_{\gamma,y|z}(1|z) + p_{\gamma,y|z}(2|z)\}$; $q^*_{\theta,y|z}(y|z) = \frac{p_{\gamma,y|z}(y|z)}{p_{\gamma,y|z}(1|z)+p_{\gamma,y|z}(2|z)}\eta(\theta;z)$ for $y = 1,2$ and $q^*_{\theta,y|z}(3|z) = 1 - \eta(\theta;z)$ if $\theta \in \Theta_2(z;p_{\gamma,y|z}) \equiv \{\theta : \eta(\theta;z) > p_{\gamma,y|z}(1|z) + p_{\gamma,y|z}(2|z)\}$; and $q^*_{\theta,y|z}(y|z) = \frac{p_{\gamma,y|z}(y|z)}{p_{\gamma,y|z}(2|z)+p_{\gamma,y|z}(3|z)}(1-\eta(\theta;z))$ for $y = 2,3$ and $q^*_{\theta,y|z}(1|z) = \eta(\theta;z)$ if $\theta \in \Theta_2(z;p_{\gamma,y|z}) \equiv \{\theta : \eta(\theta;z) < p_{\gamma,y|z}(1|z)\}$. Concerning the score function, we have $s_\theta(y|z;p_{\gamma,y|z}) = 0$ if $\theta \in \Theta_1(z;p_{\gamma,y|z})$; $s_\theta(y|z;p_{\gamma,y|z}) = \frac{-\ln(Az)(Az)^\theta}{1-(Az)^\theta}$ for $y = 1,2$ and $s_\theta(3|z;p_{\gamma,y|z}) = \ln Az$ if $\theta \in \Theta_2(z;p_{\gamma,y|z})$; $s_\theta(1|z;p_{\gamma,y|z}) = \frac{-\ln(Az)(Az)^\theta}{1-(Az)^\theta}$ and $s_\theta(y|z;p_{\gamma,y|z}) = \ln Az$ for $y = 2,3$ if $\theta \in \Theta_3(z;p_{\gamma,y|z})$.

Next, let $\bar\gamma$ be defined so that $p_{\bar\gamma,y|z}(1|z_H) = p_{\bar\gamma,y|z}(1|z_L) + p_{\bar\gamma,y|z}(2|z_L)$. Algebraic manipulations show that if $\gamma \le \bar\gamma$, then $\Theta^*(p_\gamma) = \Theta_I(p_\gamma)$ as in (C.1) with p_0 replaced by p_γ. However, $\Theta_I(p_\gamma) = \emptyset$ for all $\gamma \in (\bar\gamma, 1]$. Yet, recalling that at any $\theta^* \in \Theta^*(p_\gamma)$ it must hold that $\mathbb{E}[s_\theta(Y|Z;p_{\gamma,y|z})] = 0$, one can show that $\mathbb{E}[s_\theta(Y|Z;p_{\gamma,y|z})] = 0$ if and only if either

$$\frac{\ln(Az_L)\big[p_{\gamma,y|z}(3|z_L)-(Az_L)^\theta\big]}{1-(Az_L)^\theta} + \frac{\ln(Az_H)\big[1-(Az_H)^\theta - p_{\gamma,y|z}(1|z_H)\big]}{1-(Az_H)^\theta} = 0,$$

or

$$\frac{\ln(Az_H)\left[p_{\gamma,y|z}(3|z_H)-(Az_H)^\theta\right]}{1-(Az_H)^\theta}+\frac{\ln(Az_L)\left[1-(Az_L)^\theta-p_{\gamma,y|z}(1|z_L)\right]}{1-(Az_L)^\theta}=0,$$

but not both. Then, $\Theta^*(p_\gamma)$ equals the singleton value of θ that solves the one that holds. $\square$

Example C.2 (Panel dynamic discrete choice)**.** Decision maker i chooses between actions $y=0$ and $y=1$ across multiple time periods, according to

$$Y_{it}=1\{X_{it}\gamma+Y_{it-1}\beta+\alpha_i+\epsilon_{it}\geq 0\},\ i=1,\ldots,n,\ t=1,\ldots,T$$

with Y_{it} their decision in period t, X_{it} a vector of observed covariates in period t, α_i an individual-specific unobserved effect that is fixed over time, and ϵ_{it} an idiosyncratic unobserved effect that varies over time. When $\beta\neq 0$, period t's choice depends on previous periods' choices, introducing state dependence. The researcher observes (Y_{it},X_{it}) for $i=1,\ldots,n$ and $t=1,\ldots,T$, but does not observe Y_{i0}, so that $\{Y_{i1},\ldots,Y_{iT}\}$ is not fully determined and the model is incomplete (Heckman, 1978, Honoré and Tamer, 2006). Nonetheless, for given $(X_{it},\alpha_i,\epsilon_{it}),\ t=1,\ldots,T$, the model constrains the possible values that $(Y_{it},t=1,\ldots,T)$ can take. For example, with $T=2$, one has:

$$Y_{i1}=\begin{cases}1\{X_{i1}^\top\gamma+\alpha_i+\epsilon_{i1}\geq 0\} & \text{if } Y_{i0}=0,\\ 1\{X_{i1}^\top\gamma+\beta+\alpha_i+\epsilon_{i1}\geq 0\} & \text{if } Y_{i0}=1,\end{cases}$$

$$Y_{i2}=1\{X_{i2}^\top\gamma+Y_{i1}\beta+\alpha_i+\epsilon_{i2}\geq 0\}\quad \text{if } Y_{i0}=0 \text{ or } Y_{i0}=1.$$

Denoting the unobservables as $U_{it}\equiv\alpha_i+\epsilon_{it}$, for given $\theta=(\gamma,\beta)\in\Theta$ and $x\in\mathcal{X}$, Chen and Kaido (2023) derive the measurable (e.g., Molchanov and Molinari, 2018, Example 1.5) correspondence $G(\cdot|x;\theta)$ as the set of values $(y_1,y_2)\in\{0,1\}^2$ that satisfy the above equations. The correspondence is depicted in Panel (b) of Figure C.1 as a function of (u_1,u_2) for the case that $\beta\geq 0$. Similar examples arise in nonparametric models of state dependence (e.g., Torgovitsky, 2019). In this example, if the parameter space for β is a subset of $\mathbb{R}_{++}$, Assumption 1 (b) is satisfied because then for all $\theta\in\Theta$,

$\mathcal{A}_G = \{\{(0,0)\},\{(0,1)\},\{(1,0)\},\{(1,1)\},\{(0,1),(1,1)\},\{(0,0),(1,1)\},\{(0,0),(1,0)\}\}$. Assumption 1 (c),(e) can be verified similarly to how they are verified for Example 1. □

APPENDIX D: ALLOWING $\mathrm{supp}(G(\cdot|x;\theta))$ TO DEPEND ON θ

Assumption 1-(b) requires $\mathcal{A}_G(x) = \mathrm{supp}(G(\cdot|x;\theta))$ to be invariant across $\theta \in \Theta$. This condition can be dispensed with through the use of Clarke's subdifferentials (Clarke, 1990), at the cost of a more cumbersome procedure as we show here. For simplicity, we illustrate the approach for the two player entry game with payoffs given in Example 1 and let (δ_1,δ_2) belong to parameter space $[\underline{\delta},\bar{\delta}]^2 \ni 0$. This is an instructive case because $\mathrm{supp}(G(\cdot|x;\theta))$ changes depending on which quadrant (δ_1,δ_2) belong to, while being invariant to θ for (δ_1,δ_2) in the interior of a quadrant, and for $\delta_1 \cdot \delta_2 = 0$ the region of multiplicity in Figure 1 disappears. When $\mathrm{sign}(\delta_1 \cdot \delta_2) < 0$, a PSNE does not exist for certain values of θ (see, e.g. Tamer, 2003, Figure 2) and following Beresteanu et al. (2011, Appendix D) we let $G(\cdot|\theta) = \mathcal{Y}$ for these values of θ. Throughout, we assume that Θ only includes values for $\mathrm{Corr}(U_1,U_2)$ bounded away from ± 1.

The value function is $V(\theta|x) = \sup_{q \in \mathfrak{q}_{\theta,x}} \sum_{y\in\mathcal{Y}} p_{0,y|x}(y|x) \ln q(y)$, with $\mathfrak{q}_{\theta,x} = \mathfrak{q}^0_{\theta,x}$ if $\delta_1 \cdot \delta_2 = 0$, $\mathfrak{q}_{\theta,x} = \mathfrak{q}^{\mathrm{I}}_{\theta,x}$ if $\delta_1 > 0, \delta_2 > 0$, $\mathfrak{q}_{\theta,x} = \mathfrak{q}^{\mathrm{II}}_{\theta,x}$ if $\mathrm{sign}(\delta_1 \cdot \delta_2) < 0$, and $\mathfrak{q}_{\theta,x} = \mathfrak{q}^{\mathrm{III}}_{\theta,x}$ if $\delta_1 < 0, \delta_2 < 0$. The set $\mathfrak{q}^{\mathrm{III}}_{\theta,x}$ corresponds to the one in (3.6), while

$$
\begin{aligned}
\mathfrak{q}^0_{\theta,x} = \Big\{ q \in \Delta:\ & q((0,0)) = F_\theta((-\infty,-x_1\beta_1),(-\infty,-x_2\beta_2));\\
& q((1,1)) = F_\theta([-x_1\beta_1,\infty),[-x_2\beta_2,\infty));\\
& q((1,0)) = F_\theta([-x_1\beta_1,\infty),(-\infty,-x_2\beta_2));\\
& q((0,1)) = F_\theta((-\infty,-x_1\beta_1),[-x_2\beta_2,\infty)) \Big\},\\
\mathfrak{q}^{\mathrm{I}}_{\theta,x} = \Big\{ q \in \Delta:\ & q((0,0)) \geq F_\theta((-\infty,-x_1\beta_1-\delta_1),(-\infty,-x_2\beta_2))\\
& \qquad\qquad + F_\theta([-x_1\beta_1-\delta_1,-x_1\beta_1),(-\infty,-x_2\beta_2-\delta_2))\\
& q((0,0)) \leq F_\theta((-\infty,-x_1\beta_1),(-\infty,-x_2\beta_2));\\
& q((1,0)) = F_\theta([-x_1\beta_1,\infty),(-\infty,-x_2\beta_2-\delta_2));
\end{aligned}
$$

$$
\begin{aligned}
& q((0,1)) = F_\theta((-\infty, -x_1\beta_1 - \delta_1), [-x_2\beta_2, \infty))\Big\},\\
\mathfrak{q}^{\mathrm{II}}_{\theta,x} = \Big\{q \in \Delta:\ & q((0,0)) \geq F_\theta((-\infty, -x_1\beta_1), (-\infty, -x_2\beta_2 - \delta_2));\\
& q((1,1)) \geq F_\theta([-x_1\beta_1 - \delta_1, \infty), [-x_2\beta_2 - \delta_2, \infty));\\
& q((1,0)) \geq F_\theta([-x_1\beta_1, \infty), (-\infty, -x_2\beta_2 - \delta_2))\\
& q((0,1)) \geq F_\theta((-\infty, -x_1\beta_1 - \delta_1), (-x_2\beta_2, \infty))\Big\},
\end{aligned}
$$

with Δ the unit simplex in $\mathbb{R}^4$. We assume that $F_\theta((-\infty, t_1], (-\infty, t_2])$ is jointly continuous in (t_1, t_2, θ) and for all $t_1, t_2 \in \mathbb{R}$, $F_\theta((-\infty, t_1], (-\infty, t_2])$, $F_\theta([t_1, \infty), [t_2, \infty))$, $F_\theta([t_1, \infty), (-\infty, t_2])$, $F_\theta((-\infty, t_1], [t_2, \infty)) \in \mathbb{R}_{++}$. By Theorem 3.1-(i), for $\theta \in \Theta$ such that (δ_1, δ_2) is in the interior of a quadrant, $V(\theta|x)$ is continuously differentiable in θ, and provided $q^*_{\theta,y|x}(y|x) > c$, $P_0 - a.s.$ the gradient is bounded. Let V^{I} denote $V(\theta|x)$ for $\delta_1, \delta_2 > 0$, and use similar notation for the other cases and for the gradient of $\nabla V(\theta|x)$. In our entry game example, when $\delta_1 < 0, \delta_2 < 0$ we proved that ∇V^{III} equals the inner product of $p_{0,y|x}$ with the score vector in (3.24)-(3.27). A similar derivation can be carried out to obtain ∇V^{I} when $\delta_1 > 0, \delta_2 > 0$ and ∇V^{II} when $\delta_1 \cdot \delta_2 < 0$. Moreover, one can show that whenever $\delta_1 \cdot \delta_2 = 0$, the formulas for V coming from the adjacent quadrants coincide. Hence V is continuous on Θ. Because each gradient $\nabla V^{\mathrm{I}}, \nabla V^{\mathrm{II}}, \nabla V^{\mathrm{III}}$ is continuous and Θ is compact, the gradients are bounded on each quadrant. As there are finitely many quadrants, $V(\theta|x)$ is locally Lipschitz on θ (Clarke, 1990, p.25); however, $V(\theta|x)$ fails to be differentiable at $\theta : \delta_1\delta_2 = 0$. By Theorem 2.5.1 in Clarke (1990, p.63), for T any set of Lebesgue mesaure zero in $\mathbb{R}^{d_\theta}$, the Clarke *generalized gradient* of V at θ is

$$
\partial V^\circ(\theta|x) = \mathrm{conv}\left\{\lim \nabla V(\theta_i) : \theta_i \to \theta, \theta_i \notin T, \theta_i : \delta_{i1}\delta_{i2} \neq 0\right\}.
$$

Let $\overline{\nabla V}^{\mathrm{J}},\ \mathrm{J} \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ denote the continuous extensions of ∇V^{J} to the closure of the quadrant associated with $\mathfrak{q}^{\mathrm{J}}_{\theta,x}$, denoted Θ^{J}. Then

$$\partial V^{\circ}(\theta|x) = \begin{cases} \{\nabla V^{\mathrm{J}}(\theta|x)\}, & \theta \in \Theta^{\mathrm{J}}, \\ \mathrm{conv}\{\overline{\nabla V}^{\mathrm{I}}(\theta|x), \overline{\nabla V}^{\mathrm{II}}(\theta|x)\}, & \delta_1 = 0, \delta_2 > 0, \text{ or } \delta_1 > 0, \delta_2 = 0, \\ \mathrm{conv}\{\overline{\nabla V}^{\mathrm{II}}(\theta|x), \overline{\nabla V}^{\mathrm{III}}(\theta|x)\}, & \delta_1 = 0, \delta_2 < 0, \text{ or } \delta_1 < 0, \delta_2 = 0, \\ \mathrm{conv}\{\overline{\nabla V}^{\mathrm{I}}(\theta|x), \overline{\nabla V}^{\mathrm{II}}(\theta|x), \overline{\nabla V}^{\mathrm{III}}(\theta|x)\}, & \delta_1 = \delta_2 = 0. \end{cases}$$

Next, Theorem 3.1-(ii) shows that $\mathbb{E}[s_\theta(Y|X;p_{0,y|x})] = 0$ for all $\theta \in \Theta^*(p_0)$. Correspondingly, we first note that when vectors $\theta : \delta_1 \cdot \delta_2 = 0$ belong to Θ, $\partial V^{\circ}(\theta)$ is set valued and no longer a singleton. Correspondingly, we use the Aumann expectation of $\partial V^{\circ}(\theta|X)$ (Molchanov and Molinari, 2018, Chapter 3) to collect all the possible values of the expected score. Armed with this set-valued expectation, we invoke Proposition 2.3.2 in Clarke (1990, p.38), by which if $L(\theta)$ attains a local maximum at θ^*, then

$$0 \in \mathbb{E}\left[\partial^{\circ} V(\theta^*|X)\right].$$

For each $\theta \in \Theta$, by Clarke (1990, Proposition 2.1.2) and Molchanov and Molinari (2018, Theorem 3.11), the set $\mathbb{E}\left[\partial^{\circ} V(\theta|X)\right]$ is closed and convex P_0-a.s., hence we can represent it using its support function. Moreover, its Clarke subdifferential is equal to the gradient for θ such that $V(\theta|X)$ is differentiable P_0-a.s. As the only non-differentiability points are $\theta : \delta_1\delta_2 = 0$, and this set of values is known, we can build a confidence set for θ that dispenses with Assumption 1-(b). Fix $\epsilon > 0$, say $\epsilon = 10^{-6}$, and let

$$CS_n = \left\{\theta \in \Theta^c_\epsilon : T_n(\theta) \le c_{d_\theta,\alpha}\right\} \cup \left\{\theta \in \Theta_\epsilon : \sqrt{n}\mathrm{d}_H(0, \hat{\mathbb{E}}_n\left[\partial^{\circ} V(\theta|X_i)\right]) \le c_{H,\alpha}(\theta)\right\}, \quad \text{(D.1)}$$

with, for any nonempty and compact sets A, B, $\mathrm{d}_H(A,B) = \sup_{a\in A}\inf_{b\in B}\|a-b\|$, and $\Theta_\epsilon = \{\theta \in \Theta : |\delta_1\delta_2| < \epsilon\}$, $\Theta^c_\epsilon = \Theta \setminus \Theta_\epsilon$, and $c_{H,\alpha}(\theta)$ the $1-\alpha$ quantile of the limit distribution of $\mathrm{d}_H(\mathbb{E}\left[\partial^{\circ} V(\theta|X)\right], \hat{\mathbb{E}}_n\left[\partial^{\circ} V(\theta|X_i)\right])$ (using that $\mathrm{d}_H(0, \hat{\mathbb{E}}_n\left[\partial^{\circ} V(\theta|X_i)\right]) \le \mathrm{d}_H(\mathbb{E}\left[\partial^{\circ} V(\theta|X)\right], \hat{\mathbb{E}}_n\left[\partial^{\circ} V(\theta|X_i)\right])$ under the null that $0 \in \mathbb{E}\left[\partial^{\circ} V(\theta|X)\right]$). If $\partial^{\circ} V(\theta|X_i)$

were observed, we could adapt Beresteanu and Molinari (2008)'s results and consistent bootstrap procedure to estimate the critical values of the limit distribution, to establish asymptotic validity of CS_n in (D.1). However, in our application $\partial^\circ V(\theta|X_i)$ depends on $p_{0,y|x}$, which is unknown and nonparametrically estimated in first stage. Semenova (2023) and Liu and Molinari (2025) derive the limit distribution of support function processes similar to the one associated with $\mathrm{d}_H(\mathbb{E}\left[\partial^\circ V(\theta|X)\right], \hat{\mathbb{E}}_n\left[\partial^\circ V(\theta|X_i)\right])$, accounting for first-step nonparametric estimation, and they put forward consistent bootstrap procedures to estimate the critical values. Due to space constraints and the fact that our main proposal trades computationally tractability for the stronger Assumption 1-(b), we leave to future research extending their results to our context.

APPENDIX: REFERENCES

BARSEGHYAN, LEVON, MAURA COUGHLIN, FRANCESCA MOLINARI, AND JOSHUA C. TEITELBAUM (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89, 2015–2048. [20, 21]

BERESTEANU, ARIE, ILYA MOLCHANOV, AND FRANCESCA MOLINARI (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79, 1785–1821. [24]

BERESTEANU, ARIE AND FRANCESCA MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [27]

CHEN, SHUOWEN AND HIROAKI KAIDO (2023): "Robust Tests of Model Incompleteness in the Presence of Nuisance Parameters," available at https://arxiv.org/abs/2208.11281. [23]

CHEN, XIAOHONG AND TIMOTHY M. CHRISTENSEN (2015): "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions," *Journal of Econometrics*, 188, 447–465. [19]

CHEN, XIAOHONG, TIMOTHY M. CHRISTENSEN, AND ELIE TAMER (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965–2018. [20, 21]

CHEN, XIAOHONG, OLIVER LINTON, AND INGRID VAN KEILEGOM (2003): "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608. [15, 16, 17]

CLARKE, FRANK H. (1990): *Optimization and Nonsmooth Analysis*, Philadelphia: Society for Industrial and Applied Mathematics, 2nd ed. [24, 25, 26]

CSISZÁR, IMRE AND JÁNOS KÖRNER (2011): *Information theory: coding theorems for discrete memoryless systems*, Cambridge: Cambridge University Pr., 2nd ed. ed. [5]

GALICHON, ALFRED AND MARC HENRY (2011): "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298. [3]

GAUVIN, JACQUES AND ROBERT JANIN (1990): "Directional derivative of the value function in parametric optimization," *Annals of Operations Research*, 27, 237–252. [6]

HECKMAN, JAMES J. (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence," *Annales de l'inséé*, 227–269. [23]

HONORÉ, BO E. AND ELIE TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. [23]

KAIDO, HIROAKI AND FRANCESCA MOLINARI (2024): "Information Based Inference in Models with Set-Valued Predictions and Misspecification," available at https://arxiv.org/abs/2401.11046v1. [6]

LIU, YIQI AND FRANCESCA MOLINARI (2025): "Inference for an Algorithmic Fariness-Accuracy Frontier," available at https://arxiv.org/abs/2402.08879. [27]

LUO, YE, KIRILL PONOMAREV, AND HAI WANG (2025): "Selecting Inequalities for Sharp Identification in Models with Set-Valued Predictions," . [3]

MOLCHANOV, ILYA AND FRANCESCA MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge University Press, Cambridge UK. [1, 3, 23, 26]

POWELL, JAMES L., JAMES H. STOCK, AND THOMAS M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430. [18]

ROCKAFELLAR, R. T. (1984): *Directional differentiability of the optimal value function in a nonlinear programming problem*, Berlin, Heidelberg: Springer Berlin Heidelberg, 213–226. [6]

ROCKAFELLAR, R. TYRRELL AND ROGER J.-B. WETS (2005): *Variational Analysis, Second Edition*, Springer-Verlag, Berlin. [7]

SEMENOVA, VIRA (2023): "Debiased machine learning of set-identified linear models," *Journal of Econometrics*, 235, 1725–1746. [27]

TAMER, ELIE (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165. [24]

TORGOVITSKY, ALEXANDER (2019): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87, 1475–1505. [23]

VAN DER VAART, A. W. AND JON A. WELLNER (1996): "Weak Convergence and Empirical Processes, With Applications to Statistics," *Springer Series in Statistics*. [9, 15, 17]