Selecting Inequalities for Sharp Identification in Models with Set-Valued Predictions*

Ye Luo,[1] Kirill Ponomarev,[2] Hai Wang[3]

[1] HKU Business School
[2] University of Chicago
[3] Singapore Management University

This version: January 28, 2026

Abstract

In many partially identified econometric models, sharp identified sets can be generically characterized using specific moment inequalities known as Artstein's inequalities. Although such a characterization is theoretically appealing, the resulting collection of inequalities typically includes many redundant elements, which do not carry additional identifying information but make the analysis computationally intractable. In this paper, we characterize the smallest possible collection of moment inequalities that suffices for sharpness and provide an efficient algorithm to obtain such inequalities in practice. As a result, we obtain tractable characterizations of sharp identified sets in several well-studied settings. In situations when the smallest collection of inequalities is still infeasible, we discuss additional modeling assumptions that further simplify computation. We apply the results to the models of static and dynamic games, potential outcomes, discrete choice, network formation, selectively observed data, and ascending auctions, and demonstrate in simulations that the proposed method substantially improves upon informal inequality selection.

*Corresponding author: Kirill Ponomarev, ponomkirill@gmail.com. We thank Victor Chernozhukov, Andrew Chesher, Alfred Galichon, Hiro Kaido, Conroy Lau, Rosa Matzkin, Eric Mbakop, Francesca Molinari, Amedeo Odoni, Adam Rosen, Thomas Russell, Andres Santos, Azeem Shaikh, Xiaoxia Shi, Alex Torgovitsky, and Zhaoyang Xu for helpful comments and discussions. We thank Adam Drozynski for excellent research assistance.

1 Introduction

Many econometric models have the following structure: Given covariates $X \in \mathcal{X}$, latent variables $U \in \mathcal{U}$, and parameters $\theta \in \Theta$, the model produces a set $G(U, X; \theta) \subseteq \mathcal{Y}$ of possible values for the outcome $Y \in \mathcal{Y}$. The researcher does not observe $G(U, X; \theta)$ directly, but postulates that $Y \in G(U, X; \theta_0)$, almost surely, for some $\theta_0 \in \Theta$. The mechanism that selects a single value Y from the set $G(U, X; \theta_0)$ may be somehow restricted or left completely unspecified.[1] Examples of such settings include static and dynamic entry games (e.g., Tamer, 2003; Ciliberto and Tamer, 2009; Berry and Compiani, 2020; Gu et al., 2022); network formation models (e.g., Miyauchi, 2016; De Paula et al., 2018; Sheng, 2020; Gualdani, 2021); English auctions (e.g., Haile and Tamer, 2003; Aradillas-López et al., 2013); models with missing or interval data (e.g., Manski, 1994, 2003; Beresteanu et al., 2011); potential outcome models (e.g., Heckman et al., 1997; Manski and Pepper, 2000, 2009; Beresteanu et al., 2012; Russell, 2021); and discrete choice models with endogeneity (e.g., Chesher et al., 2013; Chesher and Rosen, 2017; Torgovitsky, 2019; Tebaldi et al., 2019) or unobserved or counterfactual choice sets (e.g., Manski, 2007; Barseghyan et al., 2021).

Sharp identified sets in such models can be generally characterized as follows. By assumption, $Y \in G(U, X; \theta_0)$, almost surely, so $\{G(U, X; \theta_0) \subseteq A\}$ implies $\{Y \in A\}$, for any measurable set $A \subseteq \mathcal{Y}$. Thus, at $\theta = \theta_0$, the inequalities

$$P(Y \in A \mid X = x) \geqslant P(G(U, X; \theta) \subseteq A \mid X = x; \theta) \tag{1}$$

must hold for all $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$. Therefore, a natural identified set for θ is

$$\Theta_0 = \{\theta \in \Theta : (1) \text{ holds for all } A \subseteq \mathcal{Y}, x \in \mathcal{X}\}. \tag{2}$$

The results of Artstein (1983) and Theorem 2.33 in Molchanov and Molinari (2018) imply that the inequalities in (1) hold if and only if $Y \in G(U, X; \theta)$, almost surely. Therefore, assuming the parameter space Θ captures all other restrictions imposed on the model, the identified set Θ_0 is sharp.

The characterization in (2) is often impractical since the total number of Artstein's inequalities may be very large. In such settings, it is customary to select a smaller collection of inequalities based on intuition or experience and proceed with an outer set for Θ_0. This approach has two important drawbacks: First, omitting an important inequality may lead

[1] In some examples, the set-valued predictions naturally arise in the space of latent variables, rather than the outcome space. Specifically, given Y, X, and θ, the model produces a set $G(Y, X; \theta)$ such that $U \in G(Y, X; \theta_0)$ for some $\theta_0 \in \Theta_0$. As discussed in Chesher and Rosen (2017), the two approaches are equivalent. Our analysis applies in both settings.

to a substantial loss of identifying information; Second, having outer identified sets that are very narrow may be a symptom of "identification by misspecification" and potentially lead to misleading conclusions (see Kédagni et al., 2020).

At the same time, examples suggest that many inequalities in (2) may be redundant, in the sense that omitting them does not change the resulting identified set. By finding and removing such inequalities, it is often possible to keep the analysis tractable while avoiding information loss and mitigating misspecification concerns. This paper proposes a simple and computationally efficient way to do so.

To address inequality selection, we focus on core-determining classes following Galichon and Henry (2011); Chesher and Rosen (2017); Luo and Wang (2018); and Molchanov and Molinari (2018). Consider the Artstein's inequalities in (1) for a fixed $X = x$. A class of $\mathcal{C}$ of subsets of $\mathcal{Y}$ is called a *core-determining class (CDC)* if verifying (1) for all $A \in \mathcal{C}$ suffices to conclude that it holds for all $A \subseteq \mathcal{Y}$. Evidently, smaller classes $\mathcal{C}$ lead to more concise characterization of the sharp identified set. In this paper, we obtain a simple analytical characterization of the smallest possible CDC. We show that such CDC depends only on the structure of the model's correspondence $G(U, x; \theta)$ and the null sets of the underlying probability distribution and typically needs to be computed only a finite number of times. We also develop a new algorithm for computing the smallest CDC, which avoids the major computational bottleneck of checking all candidate sets for redundancy. The algorithm operates by checking the connectivity of suitable subgraphs of a bipartite graph, which represents the model's correspondence, and is output-sensitite: its' computational complexity is proportional to the size of the smallest CDC. We apply the proposed methodology to obtain tractable characterizations of sharp identified sets in several well-studied settings.

This paper contributes to the large and growing literature on econometrics with partial identification; see, e.g., Pakes et al. (2015); Molinari (2020); Chesher and Rosen (2020); and Kline et al. (2021) for detailed reviews. The key object in the identification analysis is the set $\mathcal{P}(x; \theta)$ of model-implied distributions of the outcome Y, given covariates $X = x$ and a parameter value $\theta \in \Theta$. By construction, the sharp identified set for θ_0 is given by $\Theta_0 = \{\theta \in \Theta : P_{Y|X=x} \in \mathcal{P}(x; \theta),\ x \in \mathcal{X}\text{-a.s.}\}$. Existing approaches to identification are based on obtaining tractable characterizations of the set $\mathcal{P}(x; \theta)$.

Several existing papers represent the set $\mathcal{P}(x; \theta)$ using the Artstein's inequalities in (1). Galichon and Henry (2011) discuss several methods for computing sharp identified sets in discrete games. They consider submodular optimization and optimal transport approaches, which we discuss in more detail in Section 4.3, and introduce the notion of core-determining classes. In particular, they show that if the model's correspondence is suitably monotone, there exists a CDC whose size scales linearly with the size of the outcome space. In general, however, even the smallest CDC may grow exponentially with the size of the outcome space, and it is much harder to characterize. This paper extends the results of Galichon and Henry (2011) by deriving the smallest possible CDC without any restrictions on the model's correspondence and developing an efficient algorithm to compute it in practice.

Chesher and Rosen (2017) derive analytical sufficient conditions for identifying redundant Artstein's inequalities. In this paper, we obtain a richer set of necessary and sufficient conditions for redundancy and use it to characterize the smallest possible CDC. Moreover, we provide a new algorithm to compute such CDC in practice. Bontemps and Kumar (2020) characterize the smallest CDC in a class of entry games with complete information and provide an algorithm to compute it. Our Theorem 1 and Algorithm 3 yield the same characterization in this example, but apply far more generally.

Luo and Wang (2018) also give a characterization of the smallest CDC, which they call "exact," in their Theorem 2. We improve on and extend this result in several directions. First, although Theorem 1 below leads to the same CDC as Theorem 2 in Luo and Wang (2018), when coupled with Lemmas 1 and 2, it provides a more transparent and complete characterization. These new results identify the "critical" sets, which must be included in any CDC, as well as "implicit-equality" sets, for which the corresponding Artstein's inequalities always bind. Second, Corollary 1.1 establishes that the smallest CDC depends only on the supports of the random sets $G(U, x; \theta)$, conditional on $X = x$. Since the support typically has limited dependence on parameter values and covariates, this result implies that in discrete-outcome models, the CDC only needs to be computed a finite number of times and that the conditional Artstein's inequalities can be intersected,[2] which leads to a simpler characterization of sharp identified sets in many settings. Third, Theorem 1 implies an efficient algorithm for computing the smallest CDC numerically, which remains feasible far beyond Algorithm 1 of Luo and Wang (2018). Finally, in Section 5, we extend the main results to settings in which the outcome variable has infinite support.

Other closely related papers are Beresteanu et al. (2011) and Mbakop (2023). Beresteanu et al. (2011) study discrete games under different solution concepts and characterize the set $\mathcal{P}(x; \theta)$ as the Aumann expectation of a suitably defined random set. Convexity of the Aumann expectation allows to express it via the support function and thus characterize the sharp identified set through a convex optimization problem. In turn, Mbakop (2023) studies panel discrete choice models and argues that, under certain restrictions on the distribution of unobservables, the sets $\mathcal{P}(x; \theta)$ are polytopes and the inequalities that define their facets can be computed by solving a multiple-objective linear program. The CDC approach complements these methods and, as we argue below, enables faster computation and simpler

[2]Given a collection of inequalities $\theta(x) \geqslant 0$, for all $x \in \mathcal{X}$, by "intersecting" we mean $\inf_{x \in \mathcal{X}} \theta(x) \geqslant 0$.

inference procedures in many settings.

Other related work includes Tebaldi et al. (2019) and Gu et al. (2022). The former paper studies discrete choice models with endogeneity and the latter covers general discrete-outcome models. Both papers focus on obtaining sharp bounds directly on the counterfactual of interest, $\phi(\theta_0) \in \mathbb{R}$, rather than the full vector of parameters $\theta_0 \in \Theta$. They consider counterfactuals that can be expressed as linear functions of the probabilities of regions in a suitable partition of the latent variable space. If the restrictions on the distribution of latent variables induce only a finite number of linear constraints on such probabilities, the sharp bounds on the counterfactual can be obtained using linear programming. A similar approach is taken by Russell (2021), who studies a potential outcomes model with endogenous treatment assignment. The author compares different approaches to characterizing sharp bounds on functionals of the joint distribution of potential outcomes in terms of the complexity of the resulting optimization problems. In the above settings, we show that the CDC approach leads to simpler optimization problems if the smallest CDC is manageable and the excluded exogenous variables have rich support.

The algorithm we propose is related to the problems of identifying redundant constraints in linear systems (Telgen, 1983), computing a minimal half-space representation for a special class of convex polytopes (Avis and Fukuda, 1991), and listing maximal independent sets in bipartite graphs (Tsukiyama et al., 1977), and may be of independent interest. We defer a more detailed discussion to Section 4 and Appendix B.3.

The rest of the paper is organized as follows. Section 2 presents motivating examples and provides some background. Section 3 presents novel theoretical results. Section 4 discusses computation. Section 5 provides an extension to models in which the outcomes have infinite support. Section 6 illustrates the utility of selecting inequalities. Section 7 concludes.

2 Models with Set-Valued Predictions

2.1 Motivating Examples

To outline the scope of the paper, we start with three stylized examples, all of which feature outcomes with finite support. Additional examples are considered in Appendix D, and a discussion of continuous-outcome models is deferred to Section 5.

The first example is a static entry game studied by Bresnahan and Reiss (1991); Berry (1992); Tamer (2003); Ciliberto and Tamer (2009); Beresteanu et al. (2011); and Aradillas-López (2020).

Example 1 (Static Entry Game)**.** Each of N firms, indexed by $j = 1, \dots, N$, decides whether to stay out or enter the market, $Y_j \in \{0, 1\}$. The payoff of firm j is

$$\pi_j(Y, \varepsilon_j) = Y_j(\alpha_j + \delta_j N_j(Y) + \varepsilon_j),$$

where $Y = (Y_1, \dots, Y_N) \in \{0, 1\}^N$ is the outcome vector, $N_j(Y)$ is number of entrants except j, $U = (\varepsilon_1, \dots, \varepsilon_N) \in \mathbb{R}^N$ are payoff components unobserved to the researcher, and $(\alpha_j, \delta_j)_{j=1}^N \in \mathbb{R}^{2N}$ are payoff parameters. The joint distribution of latent variables U, denoted $F(\cdot; \gamma)$, is assumed to be known up to a finite-dimensional parameter $\gamma \in \mathbb{R}^{d_\gamma}$. Exogenous covariates X can be accommodated by letting $(\alpha_j, \delta_j, \gamma) = (\alpha_j(X), \delta_j(X), \gamma(X))$, but are omitted here for simplicity. The firms have complete information and play a pure-strategy Nash Equilibrium. The researcher observes $Y \in \{0, 1\}^N$ and wants to learn about features of $\theta = ((\alpha_j, \delta_j)_{j=1}^N, \gamma)$. Given U and θ, the model produces a set of predictions for Y corresponding to the set of pure-strategy Nash Equilibria:

$$G(U; \theta) = \{y \in \{0, 1\}^N : y_j = \mathbf{1}(\alpha_j + \delta_j N_j(y) + \varepsilon_j \geqslant 0), \text{ for all } j = 1, \dots, N\}.$$

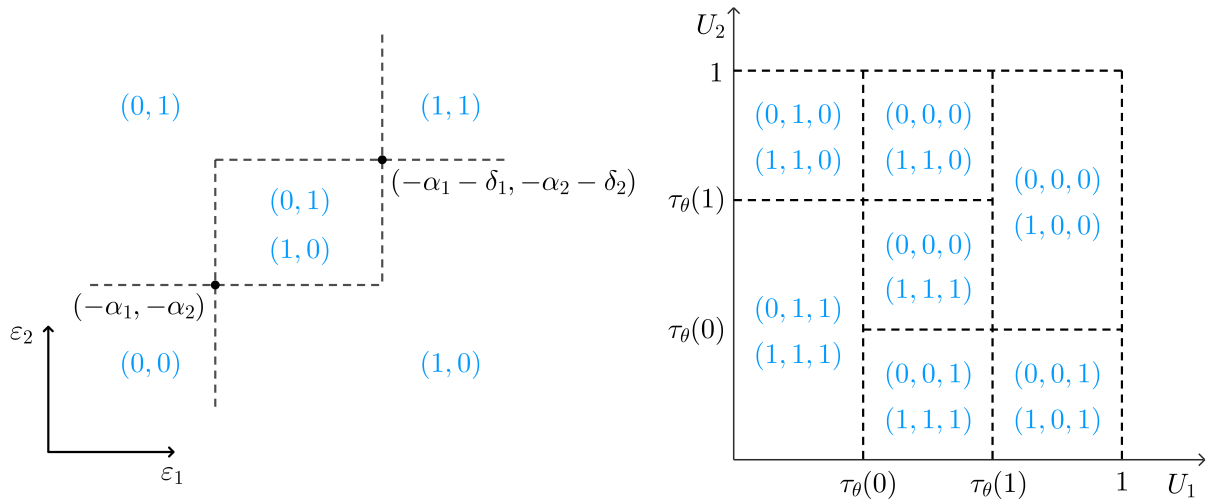
Figure 1 illustrates possible realizations of $G(U; \theta)$ when $N = 2$ and $\delta_j < 0$ for $j = 1, 2$. Dashed lines outline the partition of the latent variable space that corresponds to possible realizations of $G(U; \theta)$, highlighted in blue. ■

The next example is a simple dynamic model adapted from Berry and Compiani (2020).

Example 2 (Dynamic Monopoly Entry)**.** In time period $t = 1, \dots, T$, a firm decides to stay out of or enter the market, $A_t \in \{0, 1\}$. The per-period profit is

$$\pi(X_t, A_t, \varepsilon_t) = \begin{cases} \bar{\pi} - \varepsilon_t & \text{if } X_t = 1, A_t = 1; \\ \bar{\pi} - \varepsilon_t - \gamma & \text{if } X_t = 0, A_t = 1; \\ 0 & \text{otherwise,} \end{cases}$$

where $X_t \in \{0, 1\}$ indicates whether the firm was active in period $t - 1$, $\varepsilon_t \in \mathbb{R}$ is the variation in fixed costs, observed by the firm, and $(\bar{\pi}, \gamma)$ are the corresponding fixed profit and sunk costs of entering the market. Suppose that $\varepsilon_t = \rho\varepsilon_{t-1} + \sqrt{1 - \rho^2} v_t$ for some $\rho < 1$, and v_t are i.i.d. $N(0, 1)$. As in the preceding example, the parameters $\bar{\pi}$, γ, and ρ may depend on exogenous covariates, omitted here for simplicity. The researcher observes $Y = (X_1, A_1, \dots, A_T) \in \{0, 1\}^{T+1}$.



(a) Static entry game from Example 1 with $N = 2$ and $\delta_j < 0$ for $j = 1, 2$.

(b) Dynamic model from Example 2 with $T = 2$. Outcomes are labeled (X_1, A_1, A_2).

Figure 1: Set-valued predictions in Examples 1 and 2.

The Bellman equation for the firm's problem is

$$V(X_t, \varepsilon_t) = \max_{A_t \in \{0,1\}} \Big(\pi(X_t, A_t, \varepsilon_t) + \delta\, \mathbb{E}[V(X_{t+1}, \varepsilon_{t+1}) \,|\, A_t, X_t, \varepsilon_t] \Big),$$

where $\delta \in (0, 1)$ denotes the discount factor, which is assumed known. Under standard conditions, there is a unique stationary solution, $A_t = \mathbf{1}(U_t \leqslant \tau_\theta(X_t))$, where U_t is the quantile transformation of ε_t, and τ is an increasing function of X_t known up to $\theta = (\bar{\pi}, \gamma, \rho)$.

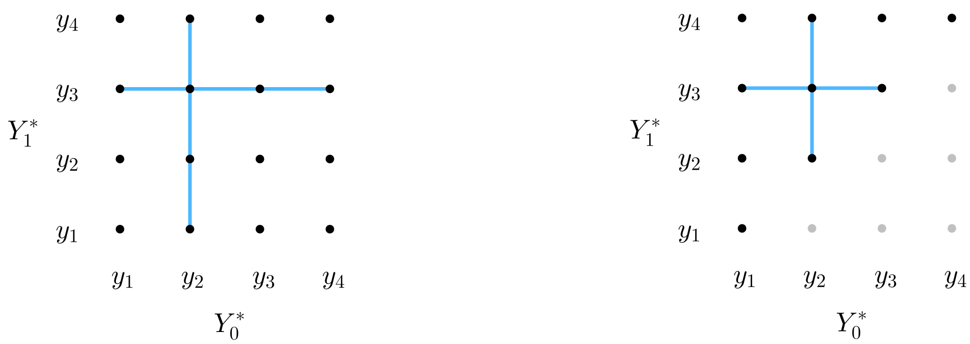
Note that X_1 is endogenous and its data-generating process is left unspecified. One way to proceed is to treat X_1 as part of the outcome vector $Y = (X_1, A_1, \ldots, A_T)$. Then, given $U = (U_1, \ldots, U_T)$ and θ, the model produces a set of possible values for Y given by

$$G(U; \theta) = \{(x_1, a_1, \ldots, a_T) : a_t = \mathbf{1}(U_t \leqslant \tau_\theta(x_t)) \text{ for } t = 1, \ldots, T\}.$$

Figure 1 illustrates possible realizations of $G(U; \theta)$ for $T = 2$. Dashed lines outline the partition of the latent variable space that corresponds to the possible realizations of $G(U; \theta)$, highlighted in blue. ■

The final example is a potential outcomes model studied in Balke and Pearl (1997), Heckman et al. (1997), Heckman and Vytlacil (2007), Beresteanu et al. (2012); Russell (2021), and Bai et al. (2024), among many others.

Example 3 (Potential Outcomes Models)**.** Let $D \in \mathcal{D}$ denote the treatment assignment,



(a) No restrictions on outcome response.

(b) Increasing outcome response, $Y_1^* \geqslant Y_0^*$.

Figure 2: Set-valued predictions Example 3 with $|\mathcal{D}| = 2$ and $|\mathcal{Y}| = 4$.

$Y^* = (Y_d^*)_{d\in\mathcal{D}} \in \mathcal{Y}^{|\mathcal{D}|}$ — potential outcomes, $Y = Y_D^*$ — observed outcome, and $Z \in \mathcal{Z}$ — instrumental variables. Suppose Y^* and Z are statistically independent and the outcome response function $d \mapsto Y_d^*$ satisfies additional restrictions summarized by $Y^* \in \mathcal{Y}^*$ for some known set $\mathcal{Y}^* \subseteq \mathcal{Y}^{|\mathcal{D}|}$ (e.g., monotonicity, partial monotonicity, concavity, etc.). Suppose the sets $\mathcal{D}$ and $\mathcal{Y}$ are finite, and $\mathcal{Z}$ is arbitrary. The primitive parameter of interest is the joint distribution of potential outcomes, $\theta = \{P(Y^* = y^*)\}_{y^*\in\mathcal{Y}^{|\mathcal{D}|}}$.

In this example, it is more convenient to construct the set-valued prediction for the latent variables Y^* given observables (Y, D, Z). If $D = d$, then $Y_d^* = Y$, but the only information available about $Y_{\tilde{d}}^*$ for $\tilde{d} \neq d$ is that $Y_{\tilde{d}} \in \mathcal{Y}$ and $Y^* \in \mathcal{Y}^*$. Thus, the set-valued prediction for Y^* can be written as:

$$G(Y, D) = B_D(Y) \cap \mathcal{S}_{Y^*},$$

where $B_d(Y) = (\mathcal{Y} \times \cdots \times \{Y\} \times \ldots \mathcal{Y})$ with $\{Y\}$ in the d-th component. Notice that Z does not affect $G(Y, D)$ in any way. Figure 2 illustrates two possible realizations of $G(Y, D)$ with $D \in \{0, 1\}$ and $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$. The vertical blue line corresponds to $G(y_2, 0)$ and the horizontal blue line to $G(y_3, 1)$. In Panel (a), $\mathcal{S}_{Y^*} = \mathcal{Y}^2$ and and in Panel (b), $\mathcal{S}_{Y^*} = \{(y, y') \in \mathcal{Y}^2 : y \leqslant y'\}$. ■

2.2 Background: Random Sets and Artstein's Inequalities

In the above examples, the set-valued prediction of the model depends on a realization of some random variables, so it is a random set. Identification in such settings can naturally be studied using tools from the theory of random sets. We briefly introduce the necessary concepts and refer the reader to Molchanov and Molinari (2018) for a textbook treatment.

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$ denote the outcome variables, $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ — observed covariates, and $U \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$ — latent variables. All random variables are defined on a common, complete probability space $(\Omega, \mathcal{F}, P)$. Let $\mathfrak{C}$ denote the class of all closed subsets of $\mathcal{Y}$, $\mathcal{B}$ — the Borel sigma-field on $\mathcal{Y}$, and $\mathcal{M}$ — the set of all probability measures on $(\mathcal{Y}, \mathcal{B})$.

Suppose the econometric model is characterized by a parameter vector $\theta \in \Theta$, which may be infinite-dimensional, and a correspondence $G(\cdot, \cdot;\, \theta) : \mathcal{U} \times \mathcal{X} \rightrightarrows \mathcal{Y}$, which delivers a set-valued prediction for the outcomes. We assume that, for each $\theta \in \Theta$, the correspondence $G(\cdot, \cdot;\, \theta)$ is measurable in the sense that $\{\omega \in \Omega : G(U(\omega), X(\omega); \theta) \subseteq A\} \in \mathcal{F}$, for all $A \in \mathfrak{C}$. We further assume that $G(U, X; \theta)$ is non-empty and closed, P-almost surely, for all $\theta \in \Theta$. Such a correspondence defines a *random closed set*. The distribution of a random closed set can be described by its' *containment functional*, defined, for all $A \in \mathfrak{C}$,[3] as

$$C_{G(U,X;\theta)}(A) = P(G(U, X; \theta) \subseteq A),$$

Any random variable Y, satisfying $P(Y \in G(U, X; \theta)) = 1$ is called a *selection* of $G(U, X; \theta)$. The set of distributions of all selections is called the *core*. Artstein (1983) showed that the core consists of all probability distributions that dominate the containment functional on closed sets, that is

$$\text{Core}(G(X, U; \theta)) = \{\mu \in \mathcal{M} : \mu(A) \geqslant C_{G(U,X;\theta)}(A),\ \text{for all } A \in \mathfrak{C}\}.$$

To characterize the core in practice, it may suffice to consider a smaller class of sets.

Definition 2.1 (Core-Determining Class)**.** *For any class of sets $\mathcal{C} \subseteq \mathfrak{C}$, denote*

$$\mathcal{M}(\mathcal{C}) = \{\mu \in \mathcal{M} : \mu(A) \geqslant C_{G(U,X;\theta)}(A),\ \textit{for all } A \in \mathcal{C}\}.$$

A class $\mathcal{C} \subseteq \mathfrak{C}$ is core-determining if $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathfrak{C})$.

We will distinguish two special types of sets.

Definition 2.2 (Critical and Implicit-Equality Sets)**.** *A set $A \in \mathfrak{C}$ is critical if $\mathcal{M}(\mathfrak{C} \backslash A) \neq \mathcal{M}(\mathfrak{C})$. A set $A \in \mathfrak{C} \backslash \{\mathcal{Y}, \varnothing\}$ is an implicit-equality set if $\mu(A) = C_{G(U,X;\theta)}(A)$, for all $\mu \in$ Core$(G(U, X; \theta))$.*

Any core-determining class must contain all critical sets and ensure that all implicit-equality constraints hold. To illustrate these definitions, suppose the outcome space $\mathcal{Y}$ is

[3] It suffices to specify the containment functional on all closed sets, but it can be extended and remains well-defined on all Borel sets. Equivalently, the distribution of a closed random set may be characterized by the *capacity functional*, $T_{G(U,X;\theta)}(A) = P(G(U, X; \theta) \cap A \neq \varnothing)$, which satisfies $T_{G(U,X;\theta)}(A) = 1 - C_{G(U,X;\theta)}(A^c)$. The same comment applies to the core of a random set defined ahead. See Sections 1.2–1.3 of Molchanov and Molinari (2018) for the details.

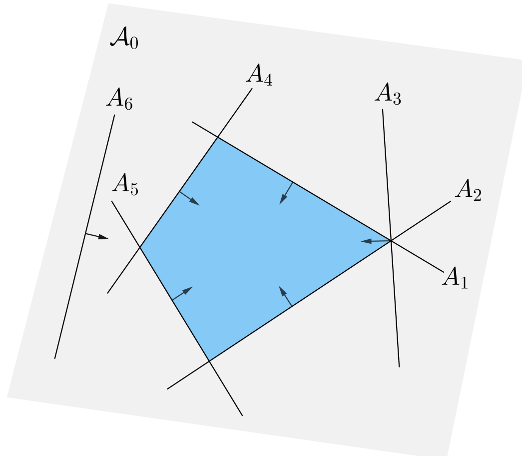


Figure 3: Stylized illustration: The core of a random set.

finite. Then, $\mathcal{M}$ is a simplex in $\mathbb{R}^{|\mathcal{Y}|}$ and $\mathrm{Core}(G(U,X;\theta))$ is a compact polyhedron, such as the one depicted in Figure 3. Here, $\mathcal{A}_0$ contains all implicit-equality sets, and the gray shaded region depicts the set $\{\mu \in \mathcal{M} : \mu(A) = C_{G(U,X;\theta)}(A),\ \text{for all } A \in \mathcal{A}_0\}$. The straight lines correspond to Artstein's inequalities with arrows indicating the directions in which they are satisfied. The core is highlighted in blue. Any class of sets that includes $\mathcal{A}_0 \cup \{A_1, A_2, A_4, A_5\}$ is core-determining. The sets A_1, A_2, A_4, A_5 are critical, while the sets A_3, A_6 are not.

3 Sharp Identified Sets with Finite Outcome Spaces

Suppose the model predicts that $Y \in G(U,X;\theta_0)$, almost surely, for some $\theta_0 \in \Theta$. With the above definitions, the sharp identified set for θ_0 can be characterized as[4]

$$\Theta_0 = \{\theta \in \Theta : P_{Y\,|\,X=x}(A) \geqslant C_{G(U,\,x;\,\theta)\,|\,X=x}(A),\ \text{for all } A \in \mathcal{C}(x,\theta),\ \text{a.s. } x \in \mathcal{X}\}, \quad (3)$$

where $\mathcal{C}(x,\theta) \subseteq \mathfrak{C}$ is a core-determining class for the random set $G(U,x;\theta)$ conditional on $X = x$, and $C_{G(U,x;\theta)\,|\,X=x}(A) = P(G(U,x;\theta) \subseteq A \,|\, X = x)$ is the conditional capacity functional. In this section, we characterize the smallest possible core-determining class $\mathcal{C}^*(x;\theta)$, clarify how it depends on x and θ, and obtain sharp identified sets in several applications. Until Section 5, we focus on settings with a finite outcome space, so $\mathfrak{C} = \mathcal{B} = 2^{\mathcal{Y}}$.

[4]The equivalence between the unconditional and conditional Artstein's inequalities follows from Theorem 2.33 in Molchanov and Molinari (2018).

3.1 Graph Representation

Fix some $x \in \mathcal{X}$ and $\theta \in \Theta$. Let $\mathcal{S}(x,\theta) = \{G_1, \ldots, G_K\}$ denote the support of $G(U,x;\theta)$ conditional on $X = x$, i.e., the set of sets $G_k \subseteq \mathcal{Y}$ such that $P(G(U,x;\theta) = G_k \,|\, X = x) > 0$. Partition the latent variable space $\mathcal{U}$ accordingly, $u_k = \{u \in \mathcal{U} : G(u,x;\theta) = G_k\}$, and denote $\mathcal{U}(x,\theta) = \{u_1, \ldots u_K\}$. Define a probability measure $P_{(x,\theta)}$ on $\mathcal{U}(x,\theta)$ by $P_{(x,\theta)}(u_k) = P_{U|X=x}(\{u : G(u,x;\theta) = G_k\})$. Then, the random set $G(U,x;\theta)$, conditional on $X = x$, can be viewed as a correspondence $G : (\mathcal{U}(x,\theta), 2^{\mathcal{U}(x,\theta)}, P_{(x,\theta)}) \rightrightarrows \mathcal{Y}$ between two finite spaces.

In what follows, with some abuse of notation, we denote $\mathcal{U}(x;\theta)$ by $\mathcal{U}$ and $P_{(x;\theta)}$ by P, let $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ denote the random set $G(U,x;\theta)$, conditional on $X = x$, and $C_G(\cdot)$ denote the conditional containment functional, $C_G(A) = P(G(U,x;\theta) \subseteq A \,|\, X = x)$. For each $A \subseteq \mathcal{Y}$, we denote the *lower and upper inverses* of G by

$$G^-(A) = \{u_k \in \mathcal{U} : G(u_k) \subseteq A\}; \qquad G^{-1}(A) = \{u_k \in \mathcal{U} : G(u_k) \cap A \neq \varnothing\}$$

Note that $G^-(A) \subseteq G^{-1}(A)$.

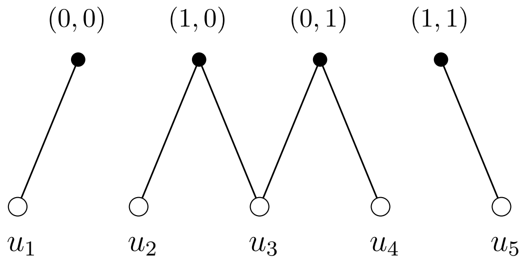
We represent the correspondence G by an undirected bipartite graph $\mathbf{B}$ with vertices $V(\mathbf{B}) = (\mathcal{Y},\mathcal{U})$ and edges $E(\mathbf{B}) = \{(u,y) \in \mathcal{U} \times \mathcal{Y} : y \in G(u)\}$. For any given x and θ, the graph $\mathbf{B}$ can be constructed either analytically or numerically, by partitioning the latent variable space as in Figure 1. Note that, although the thresholds defining the partition depend on x and θ, the graph stays the same as long as each of the regions in the partition has non-zero probability. The following examples illustrate.

Example 1 − 3 (Continued)**.** Figure 4 presents the bipartite graphs for Examples 1 − 3.

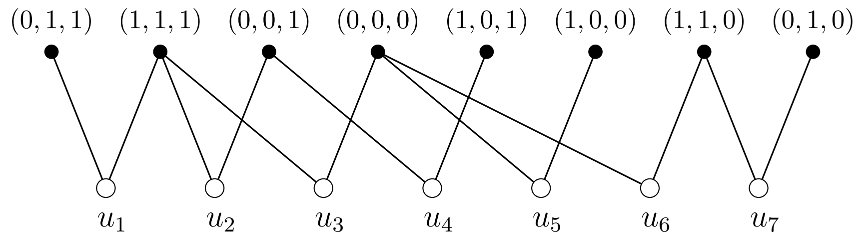
Panel (a) depicts the binary entry game with negative spillovers from Example 1. The upper part represents the outcome space $\{0,1\}^2$, and the lower part represents the partition of latent variable space illustrated in Figure 1. For example, $u_1 = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 : \varepsilon_j < -\alpha_j, j = 1,2\}$, and $u_3 = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 : -\alpha_j \leqslant \varepsilon_j < -\alpha_j - \delta_j\}$. Also, for example, $G(u_3) = \{(1,0),(0,1)\}$, $G^-(\{(1,0)\}) = u_2$, and $G^{-1}(\{(1,0),(0,1)\}) = \{u_2, u_3, u_4\}$.

Panel (b) depicts the dynamic monopoly entry model from Example 2 with $T = 2$. The upper part represents the outcome space $\{0,1\}^3$ with outcomes labeled as (x_1, a_1, a_2), and the lower part represents the partition of latent variable space illustrated in Figure 1. For example, $u_2 = \{(U_1,U_2) \in [0,1]^2 : \tau_\theta(0) < U_1 \leqslant \tau_\theta(1), U_2 \leqslant \tau_\theta(0)\}$, and $u_5 = \{(U_1,U_2) \in [0,1]^2 : U_1 > \tau_\theta(1), U_2 > \tau_\theta(0)\}$. Also, for example, $G(\{u_1, u_3\}) = \{(0,1,1),(1,1,1),(0,0,0)\}$ and $G^{-1}(\{(0,1,1),(1,1,1),(0,0,0)\}) = \{u_1, u_2, u_3, u_5, u_6\}$.

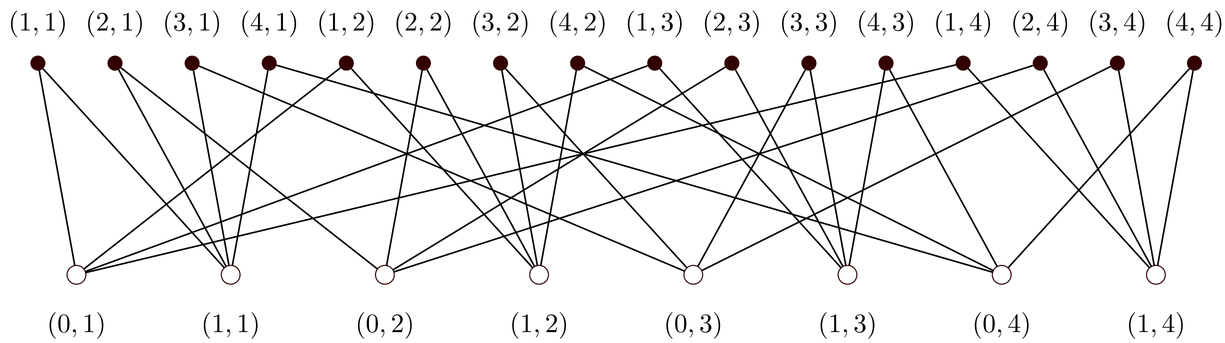
Panel (c) depicts the potential outcomes model from Example 3 with $\mathcal{D} = \{0,1\}$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and $\mathcal{S}_{Y^*} = \mathcal{Y}^2$. The upper part is $\mathcal{S}_{Y^*}$, and the lower part is $\mathcal{D} \times \mathcal{Y}$. For example, $G((0,2)) = \{(2,1),(2,2),(2,3),(2,4)\}$ corresponds to the blue vertical



(a) Entry game from Example 1 with $N = 2$ and $\delta_j < 0$ for $j = 1, 2$.



(b) Dynamic binary choice model from Example 2 with $T = 2$.



(c) Potential outcomes model from Example 3 with $D = \{0, 1\}$, $\mathcal{Y} = \{1, 2, 3, 4\}$, $\mathcal{S}_{Y^*} = \mathcal{Y}^2$.

Figure 4: Bipartite graphs in Examples 1 – 3.

line and $G((1, 3)) = \{(1, 3), (2, 3), (3, 3), (4, 3)\}$ corresponds to the blue horizontal line in Panel (a) of Figure 2. Also, for example, $G^{-}(\{(2, 1), (2, 2), (2, 3), (2, 4)\}) = \{(0, 2)\}$, and $G^{-1}(\{(2, 1), (2, 2), (2, 3), (2, 4)\}) = \{(1, 1), (0, 2), (1, 2), (1, 3), (1, 4)\}$. ■

3.2 The Structure of Redundant Inequalities

The redundancy of Artstein's inequalities can be expressed in terms of the connectivity of suitable subgraphs of the graph $\mathbf{B}$. A *subgraph of* $\mathbf{B}$ *induced by the vertices* $(V_{\mathcal{Y}}, V_{\mathcal{U}})$ is an undirected graph with vertices $(V_{\mathcal{Y}}, V_{\mathcal{U}})$ and edges $\{(u, y) \in E(\mathbf{B}) : u \in V_{\mathcal{U}}, y \in V_{\mathcal{Y}}\}$. A graph is said to be *connected* if every vertex can be reached from any other vertex through a sequence of edges.

First, suppose that for some $A \subseteq \mathcal{Y}$, there are sets $A_1, A_2 \subseteq \mathcal{Y}$ such that $A_1 \cap A_2 = \varnothing$, $A_1 \cup A_2 = A$, and $G^-(A_1 \cup A_2) = G^-(A_1) \cup G^-(A_2)$. Here, the latter condition means that $G \subseteq A_1 \cup A_2$ if and only if either $G \subseteq A_1$ or $G \subseteq A_2$, so $C_G(A_1) + C_G(A_2) = C_G(A)$. Then, summing up the inequalities $\mu(A_1) \geqslant C_G(A_1)$ and $\mu(A_2) \geqslant C_G(A_2)$, we obtain

$$\mu(A) = \mu(A_1) + \mu(A_2) \geqslant C_G(A_1) + C_G(A_2) = C_G(A), \tag{4}$$

so A is redundant given A_1 and A_2. For example, consider the graph in Panel (b) of Figure 4. Let $A_1 = \{(0,1,1),(1,1,1)\}$, $A_2 = \{(1,1,0),(0,1,0)\}$, and $A = A_1 \cup A_2$. Then, $G^-(A_1) = \{u_1\}, G^-(A_2) = \{u_7\}$, and $G^-(A) = \{u_1, u_7\}$. Thus, all of the above conditions are satisfied, and A is redundant given A_1 and A_2. Importantly, note that the subgraph induced by $(A, G^-(A))$ is disconnected.

As a special case, suppose that $A \neq G(G^-(A))$, where $G(G^-(A)) = \bigcup_{\omega \in \Omega}\{G(\omega) : G(\omega) \subseteq A\}$. That is, the set A cannot be expressed as a union of elements of the support of G. Letting $A_1 = G(G^-(A))$ and $A_2 = A \backslash A_1$, we have $G^-(A_1) = G^-(A)$ and $G^-(A_2) = \varnothing$. The inequality $\mu(A_2) \geqslant C_G(A_2) = 0$ holds trivially, so, following the argument in (4), A is redundant given A_1.[5] Consider again the graph in Panel (b) of Figure 4. Let $A = \{(0,0,1),(0,0,0),(1,0,1)\}$ and $A_1 = \{(0,0,1),(1,0,1)\} \subset A$. The set A cannot be expressed as the union of elements of the support, and $G^-(A) = G^-(A_1) = \{u_4\}$. Thus, A is redundant given A_1. As before, note that the subgraph induced by $(A, G^-(A))$ is disconnected.

Second, suppose that for some $A \subseteq \mathcal{Y}$ there are sets $A_1, A_2 \neq A$ such that $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, and $G^-(A_1) \cup G^-(A_2) = \mathcal{U}$. The latter condition means that for all $u \in \mathcal{U}$, either $G(u) \subseteq A_1$ or $G(u) \subseteq A_2$, which implies $C_G(A_1) + C_G(A_2) = 1 + C_G(A_1 \cap A_2)$. Then, adding up the inequalities $\mu(A_1) \geqslant C_G(A_1)$ and $\mu(A_2) \geqslant C_G(A_2)$, we obtain

$$1 + \mu(A) = \mu(A_1) + \mu(A_2) \geqslant C_G(A_1) + C_G(A_2) = 1 + C_G(A), \tag{5}$$

so A is redundant given A_1 and A_2. The above conditions can be equivalently stated as $A_1^c \cup A_2^c = A^c$, $A_1^c \cap A_2^c = \varnothing$, and $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \varnothing$. Returning to Panel (b) of Figure 4, let $A = \{(1,1,1),(0,0,1),(0,0,0)\}$, $A_1 = A \cup \{(0,1,1)\}$, and $A_2 = A \cup \{(1,0,1),(1,0,0),(1,1,0),(0,1,0)\}$, so that $A_1 \cap A_2 = A$. Then, $G^-(A_1) = \{u_1, u_2, u_3\}$ and $G^-(A_2) = \{u_2, u_3, u_4, u_5, u_6, u_7\}$, so that $G^-(A_1) \cup G^-(A_2) = \mathcal{U}$. Therefore, A is redundant given A_1 and A_2, as in Equation (5). In this case, the subgraph induced by $(A^c, G^{-1}(A^c))$ is disconnected.

[5] This observation implies that we may restrict attention to sets A which can be expressed as unions of elements of the support. See the errata to Beresteanu et al. (2012), Chesher and Rosen (2017), and Theorems 2.22–2.23 in Molchanov and Molinari (2018) for related arguments.

Thus, for any set $A \subseteq \mathcal{Y}$ that is redundant according to (4) or (5), the subgraph of $\mathbf{B}$ induced by either $(A, G^{-}(A))$ or by $(A^c, G^{-1}(A^c))$ is disconnected. As we show below, this simple property characterizes *all* redundant sets.

3.3 The Smallest Core-Determining Class

Following the above discussion, we say that a set $A \subseteq \mathcal{Y}$ is *self-connected* if the subgraph of $\mathbf{B}$ induced by $(A, G^{-}(A))$ is connected. Say that A is *complement-connected* if the subgraph of $\mathbf{B}$ induced by $(A^c, G^{-1}(A^c))$ is connected. Our first result characterizes the critical sets.

Lemma 1. *Let $\mathcal{U} = \{u_1, \dots, u_K\}$ and $\mathcal{Y} = \{y_1, \dots, y_S\}$ be finite sets, and $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ be a non-empty random set with a bipartite graph $\mathbf{B}$. Suppose that $\mathbf{B}$ is connected and $P(u_k) > 0$, for all $u_k \in \mathcal{U}$. A set $A \in 2^{\mathcal{Y}} \backslash \{\mathcal{Y}, \varnothing\}$ is critical if and only if it is self-connected and complement-connected.*

The proof of this result is constructive: Given a set A that is both self- and complement-connected, we construct a distribution $\mu \in \text{Core}(G)$ such that $\mu(A) = C_G(A)$ and $\mu(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$. This implies that the set $\{\mu \in \text{Core}(G) : \mu(A) = C_G(A)\}$ corresponds to one of the facets of $\text{Core}(G)$,[6] meaning that A is critical. For example, consider the set $A = \{(0,1,1), (1,1,1), (0,0,1)\}$ in Panel (b) of Figure 4. We have $G^{-}(A) = \{u_1, u_2\}$, $A^c = \{(0,0,0), (1,0,0), (1,0,0), (1,1,0), (0,1,0)\}$, and $G^{-1}(A^c) = \{u_3, u_4, u_5, u_6, u_7\}$, so A is both self- and complement-connected. Therefore, A is critical.

The assumption $P(u_k) > 0$, for all $u_k \in \mathcal{U}$, merely ensures that there are no redundant elements in $\mathcal{U}$. Any u_k with $P(u_k) = 0$ can simply be removed from $\mathcal{U}$ and $\mathbf{B}$ together with all its edges. In turn, as we show next, the assumption that $\mathbf{B}$ is connected is substantive.

Our second result characterizes the implicit-equality sets.

Lemma 2. *Let $\mathcal{U} = \{u_1, \dots, u_K\}$ and $\mathcal{Y} = \{y_1, \dots, y_S\}$ be finite sets, and $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ a non-empty random set with a bipartite graph $\mathbf{B}$. Let $\mathcal{Y} = \bigcup_{l=1}^{L} \mathcal{Y}_l$ be the finest partition of the outcome space such that $\mathcal{Y}_k \cap \mathcal{Y}_l = \varnothing$ and $G^{-1}(\mathcal{Y}_k) \cap G^{-1}(\mathcal{Y}_l) = \varnothing$, for all $k \neq l$. Then, A is an implicit-equality set if and only $A = \bigcup_{l \in L_A} \mathcal{Y}_l$ for some $L_A \subseteq \{1, \dots, L\}$.*

That is, $(\mathcal{Y}_l)_{l=1}^{L}$ are the "basic" implicit-equality sets, satisfying $\mu(\mathcal{Y}_l) = C_G(\mathcal{Y}_l)$, for all $\mu \in \text{Core}(G)$. These constraints are linearly dependent since $\sum_{l=1}^{L} \mu(\mathcal{Y}_l) = \sum_{l=1}^{L} C_G(\mathcal{Y}_l) = 1$, so any single one of them can be omitted without loss. The sets $\mathcal{Y}_l$ are easy to detect in practice: The graph $\mathbf{B}$ "breaks" into L connected components $\mathbf{B}_l$ with vertices $V(\mathbf{B}_l) = (\mathcal{Y}_l, G^{-1}(\mathcal{Y}_l))$ and edges $E(\mathbf{B}_l) = \{(u, y) \in G^{-1}(\mathcal{Y}_l) \times \mathcal{Y}_l : y \in G(u)\}$. For example, in Panel

[6] See, e.g., Theorem 8.1. in Schrijver (1998).

(a) of Figure 4, the implicit-equality sets are $\{(0,0)\}$, $\{(1,1)\}$, and $\{(1,0),(0,1)\}$. In panels (b)–(c), the graph $\mathbf{B}$ is connected, so there are no implicit-equality sets.

Combining Lemmas 1 and 2 yields our main result, characterizing the smallest CDC.

Theorem 1. *Let $\mathcal{U} = \{u_1, \ldots, u_K\}$ and $\mathcal{Y} = \{y_1, \ldots, y_S\}$ be finite sets, and $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ a non-empty random set with a bipartite graph $\mathbf{B}$ and containment functional $C_G(\cdot)$. Suppose $P(u_k) > 0$ for all $u_k \in \mathcal{U}$. The following statements hold.*

1. *Suppose $\mathbf{B}$ is connected. Let $\mathcal{C}^*$ denote the class of all critical sets, as in Lemma 1. Then,*

$$Core(G) = \{\mu \in \mathcal{M} : \mu(A) \geqslant C_G(A), \ \ \forall A \in \mathcal{C}^*\}.$$

   Moreover, $\mathcal{C}^*$ is the smallest CDC.

2. *Suppose $\mathbf{B}$ can be decomposed into connected components, $(\mathbf{B}_l)_{l=1}^L$, as in Lemma 2. Let $\mathcal{C}_l^*$ denote the class of all critical sets in $\mathbf{B}_l$, as in Lemma 1. Then,*

$$Core(G) = \{\mu \in \mathcal{M} : \mu(A) \geqslant C_G(A), \ \forall A \in \mathcal{C}_l^*, \ \mu(\mathcal{Y}_l) = C_G(\mathcal{Y}_l), \ \forall l \in \{1, \ldots, L\}\}$$

   Moreover, $\bigcup_{l=1}^L \mathcal{C}_l^* \cup \mathcal{Y}_l$ is the smallest CDC (up to removing a single arbitrary $\mathcal{Y}_l$).

This result has two key implications. For future reference, we state the first implication as a corollary. Recall the discussion in Section 3.1.

Corollary 1.1. *For any $x \in \mathcal{X}$ and $\theta \in \Theta$, let $\mathcal{S}(x;\theta)$ denote the support and $\mathcal{C}^*(x;\theta)$ denote the smallest core-determining class of the random set $G(U, x; \theta)$, conditional on $X = x$. If $\mathcal{S}(x;\theta) = \mathcal{S}(x', \theta')$ for some $\theta, \theta' \in \Theta$ and $x, x' \in \mathcal{X}$, then $\mathcal{C}^*(x;\theta) = \mathcal{C}^*(x';\theta')$.*

That is, the smallest core-determining class only depends on the support of the underlying random set. As Gu et al. (2022) point out, in discrete-outcome models, the parameter space can typically be partitioned as $\Theta = \bigcup_{m=1}^M \Theta_m$, with $\Theta_m \cap \Theta_l = \varnothing$ for $m \neq l$, so that $\mathcal{S}(x;\theta) = \mathcal{S}_m(x)$ for all $\theta \in \Theta_m$, for each $m \in \{1, \ldots, M\}$. Then, $\mathcal{C}^*(x, \theta) = \mathcal{C}_m^*(x)$ for all $\theta \in \Theta_m$, so the sharp identified set for θ can be expressed as

$$\Theta_0 = \bigcup_{m=1}^M \left\{\theta \in \Theta_m : P_{Y|X=x}(A) \geqslant C_{G(U,x;\theta)}(A), \text{ for all } A \in \mathcal{C}_m^*(x), \ x \in \mathcal{X}\right\}.$$

Additionally, it is often the case that $\mathcal{S}(x;\theta) = \mathcal{S}(x';\theta)$ for all $x, x' \in \mathcal{X}$, all $\theta \in \Theta_m$. Then, $\mathcal{C}^*(x, \theta) = \mathcal{C}_m^*$ for all $\theta \in \Theta_m$ and all $x \in \mathcal{X}$, so the sharp identified set for θ is

$$\Theta_0 = \bigcup_{m=1}^M \left\{\theta \in \Theta_m : \operatorname*{essinf}_{x \in \mathcal{X}} \left(P_{Y|X=x}(A) - C_{G(U,x;\theta)}(A)\right) \geqslant 0, \text{ for all } A \in \mathcal{C}_m^*\right\}.$$

Examples in the following section illustrate.

The second key implication of Theorem 1 is that the smallest CDC can be computed by checking the connectivity of suitable subgraphs of $\mathbf{B}$. This feature allows us to devise an algorithm that avoids the major computational bottleneck of checking all $2^{|\mathcal{Y}|} - 2$ candidate inequalities for redundancy. Further details are provided in Section 4.

3.4 Discussion and Applications

In this section, we apply Theorem 1 to characterize sharp identified sets in Examples 1–3. We show that the smallest CDC often leads to a much more tractable characterization of the sharp identified set and only needs to be computed a few times across the values of θ and X. In some settings, even the smallest CDC is too large to be practically useful, so we consider additional restrictions on the structure of the model's correspondence to simplify the analysis without losing sharpness. Examples 2 and 3 consider instrumental variables. The online appendix contains additional applications to discrete choice with endogeneity and directed network formation.

Example 1 (Continued)**.** First, suppose $\delta_j < 0$ for all j, so firms compete with each other upon entering the market.[7] For $N = 2$, the partition of the space of latent variables is illustrated in Figure 1, and the corresponding bipartite graph is in Panel (a) of Figure 4. While the regions in the partition and their corresponding probabilities change with the values of $\theta = ((\alpha_j, \delta_j)_{j=1}^N, \gamma)$, the bipartite graph remains the same as long as all $\delta_j < 0$. Therefore, the smallest CDC only needs to be computed once. The same conclusion applies when $\alpha_j(x)$ and $\delta_j(x)$ are functions of exogenous covariates, as long as $\delta_j(x) < 0$ for all $j = 1, \ldots, N$, a.s. $x \in \mathcal{X}$. Assuming that $U = (\varepsilon_1, \ldots, \varepsilon_N)$ and X are statistically independent, the sharp identified set for θ can be expressed as

$$\Theta_0 = \{\theta \in \Theta : \text{essinf}_{x \in \mathcal{X}} \left(P_{Y|X=x}(A) - C_{G(U,x;\theta)\,|\,X=x}(A)\right) \geqslant 0, \text{ for all } A \in \mathcal{C}^*\}.$$

In this model, the set of Nash Equilibria can only contain equilibria with the same number of entrants, $n \in \{0, 1, \ldots, N\}$, so the outcome space can be partitioned accordingly, $\mathcal{Y} = \bigcup_{n=0}^N \mathcal{Y}_n$, and the bipartite graph $\mathbf{B}$ breaks down into N disjoint pieces. This property dramatically reduces the CDC, because all sets of the form $A = \bigcup_{n=0}^N A_n$, where $A_n \subseteq \mathcal{Y}_n$, are redundant.[8] Table 1a summarizes the results for $N \in \{2, \ldots, 6\}$. Although the CDC is substantially smaller than the power set of the outcome space, it quickly becomes intractable.

[7]See Berry (1992) for a detailed discussion and microfoundation.

[8]This fact follows from Theorem 1 or, alternatively, Theorem 3 from Chesher and Rosen (2017) or Theorem 2.23 from Molchanov and Molinari (2018).

Next, suppose $\delta_j > 0$, which may be interpreted as that the firms are forming a coalition or a joint R&D venture. In this case, the set of Nash Equilibria only contains equilibria with different numbers of entrants. As before, whereas the relevant partition of the latent variable space and the corresponding probabilities change with θ, the bipartite graph stays the same as long as all $\delta_j > 0$ and the CDC only needs to be computed once. Table 1a summarizes the results for $N \in \{2, \ldots, 6\}$. As before, even the smallest CDC quickly becomes intractable.

If the sign of δ_j is *ex ante* unknown, the parameter space Θ can be partitioned into $M = 3^N$ regions $\Theta_1, \ldots, \Theta_M$, according to $\delta_j < 0, \delta_j = 0$, or $\delta_j > 0$ for each j, and the CDC should be computed separately for each region. For typical payoff specifications, δ_j does not depend on any exogenous characteristics x, so the support of the random set $G(U, x; \theta)$, conditional on $X = x$, does not depend on x.

The analysis can be simplified by restricting firm heterogeneity. For example, suppose that (i) there are two types of firms such that all firms within each type are identical, including the unobserved cost shifters; (ii) the profit functions depend only on the numbers of entrants of each type but not their identities.[9] Specifically, suppose the profit of firm $j \in \{1, \ldots, N\}$ of type $t \in \{1, 2\}$ takes the form

$$\pi_j^t(Y) = \begin{cases} \alpha_1 + \alpha_2(N_j^1(Y) + N_j^2(Y)) + \varepsilon_1 & t = 1; \\ \beta_1 + \beta_2 N_j^1(Y) + \beta_3 N_j^2(Y) + \varepsilon_2 & t = 2, \end{cases}$$

where $N_j^t(Y)$ is the number of entrants of type t other than firm j. Suppose $\alpha_1, \beta_2, \beta_3 < 0$ and $\beta_3 \geqslant \beta_2$. With $\beta_3 = \beta_2$, this is a direct simplification of the fully heterogeneous model discussed above. With $\beta_3 > \beta_2$, the firms compete in an asymmetric manner (e.g., type-1 firms are large and type-2 firms are small). With this payoff structure, the outcomes can be grouped together by the number of entrants of each type. Letting N^t denote the number of potential entrants of type $t \in \{1, 2\}$, the outcome space is $\tilde{\mathcal{Y}} = \{0, 1, \ldots, N^1\} \times \{0, 1, \ldots, N^2\}$, which leads to much simpler CDCs. Table 1a shows that the smallest CDC remains tractable for different compositions of firm types. Three or more types can also be accommodated. ■

Example 2 (Continued)**.** For $T = 2$, the relevant partition of the latent variable space is given in Figure 1, and the corresponding bipartite graph in Panel (b) of Figure 4. As long as $x \mapsto \tau_\theta(x)$ is strictly increasing, the structure of the bipartite graph does not depend on θ, so the smallest CDC needs to be computed only once. Let $Z \in \mathcal{Z}$ denote an excluded

[9] A version of this model with only one type leads back to Bresnahan and Reiss (1991). The model with two types was proposed by Berry and Tamer (2006) and also studied in detail by Beresteanu et al. (2008), Galichon and Henry (2011), and Luo and Wang (2018).

Heterogeneous firms					
N	2	3	4	5	6
Total	14	254	65,534	10^9	10^{19}
Smallest; $\delta_j < 0$	4	15	94	2,109	10^6
Smallest; $\delta_j > 0$	5	14	23,770	–	–
Two types of firms					
(N^1, N^2)	(1, 1)	(2, 2)	(2, 4)	(2, 7)	(6, 6)
Total	14	62	32,766	10^8	10^{14}
Smallest; $\beta_3 = \beta_2$	5	11	17	26	35
Smallest; $\beta_3 > \beta_2$	5	14	33	57	200

(a) Entry games in Example 1.

T	2	3	4	5	6	7	8	9	10
Total	30	65,534	10^9	10^{19}	10^{38}	10^{77}	10^{154}	10^{308}	10^{616}
Smallest	10	22	46	94	190	382	766	1,534	3,070

(b) Dynamic binary choice model from Example 2.

Table 1: Total number of inequalities and size of the smallest core-determining class.

Note: Symbol "–" indicates that Algorithm 3 implemented in Julia did not finish within 1 minute.

instrumental variable independent of U. Then, the sharp identified set for θ is

$$\Theta_0 = \{\theta \in \Theta : \text{essinf}_{z\in\mathcal{Z}} P(Y \in A \mid Z = z) - P(G(U;\theta) \subseteq A) \geqslant 0 \text{ for all } A \in \mathcal{C}^*\}.$$

In this example, the bipartite graph $\mathbf{B}$ that corresponds to the model's correspondence has a simple structure: Each vertex u_j has exactly two neighbors, which correspond to $x_1 \in \{0, 1\}$. As a result, while the power set of the outcome space has cardinality $2^{2^{T+1}}$, the smallest CDC grows proportionally to 2^T. Table 1b summarizes the results for $T \in \{1, \ldots, 10\}$.

In more elaborate dynamic oligopoly models, discussed by Berry and Compiani (2020), one can adopt a type-heterogeneity assumption similar to the one in Example 1 to keep the analysis tractable. The details are left for future research. ■

Example 3. (Continued) The parameter of interest is the joint distribution of potential outcomes, $\theta = P_{Y^*}$, with a known support $\mathcal{S}_{Y^*}$. Since the support of the random set $G(Y, D)$ does not depend on θ or Z, no partitioning of the parameter space is required, and the smallest CDC needs to be computed only once. Moreover, since Z is independent of Y^*,

$$\Theta_0 = \{\theta = P_{Y^*} : P_{Y^*}(A) \geqslant \text{esssup}_{z\in\mathcal{Z}} P(G(Y, D) \subseteq A \mid Z = z), \text{ for all } A \in \mathcal{C}^*\},$$

Unrestricted outcome response							
$\lvert\mathcal{D}\rvert = 2 \setminus \lvert\mathcal{Y}\rvert$	2	3	4	5	6	7	8
Total	16	512	65,534	10^7	10^{11}	10^{14}	10^{19}
Smallest	8	42	204	910	3,856	15,890	64,532
Monotone outcome response							
$\lvert\mathcal{D}\rvert \setminus \lvert\mathcal{Y}\rvert$	2	3	4	5	6	7	8
2	4	12	36	124	468	1836	7300
3	6	33	220	1,719	14,002	114,349	–
4	8	82	1,126	18,087	297,585	–	–
Monotone and concave outcome response							
$\lvert\mathcal{D}\rvert \setminus \lvert\mathcal{Y}\rvert$	2	3	4	5	6	7	8
3	4	17	81	504	3,470	25,689	194,074
4	4	17	110	973	10,106	121,755	–

Table 2: Core-determining classes in the potential outcomes model from Example 3.

Note: Symbol "–" indicates that Algorithm 3 implemented in Julia did not finish within 1 minute.

where $\mathcal{C}^*$ denotes the smallest CDC.[10]

Let us now examine the size of $\mathcal{C}^*$. First, consider the model without any restrictions on the support of Y^*. The corresponding bipartite graph (e.g., Panel (c) of Figure 4) is connected, so there are no implicit-equality sets, and all critical sets can be described analytically. Unions of elements of the support of $G(Y, D)$ are "lattice-shaped" sets $A = B_1 \times B_2 \cdots \times B_{|\mathcal{D}|}$, where each $B_d \subseteq \mathcal{Y}$ (but not necessarily singleton, as in Figure 2). If at least two of the sets B_d are strict subsets of $\mathcal{Y}$, any configuration of the remaining $|\mathcal{D}| - 2$ sets $B_{d'}$ leads to a critical set A. If $B_d \subset \mathcal{Y}$ for some d, and $B_{d'} = \mathcal{Y}$ for all $d' \neq d$, the corresponding Artstein's inequalities restrict only the marginal distribution of the Y^*_d, so it suffices to consider singleton B_d. Thus, the total number of critical sets is $\sum_{k=2}^{|\mathcal{D}|} \binom{|\mathcal{D}|}{k} (2^{|\mathcal{Y}|} - 2)^k + |\mathcal{Y}||\mathcal{D}|$. Panel (a) of Table 2 provides some examples with $|\mathcal{D}| = 2$.

Next, consider imposing constraints on the outcome response function $d \mapsto Y^*_d$. Suppose $\mathcal{D} = \{d_1, \ldots, d_{|\mathcal{D}|}\}$ is totally ordered. Then, for example, setting $\mathcal{S}^I_{Y^*} = \{y^* \in \mathcal{Y}^{|\mathcal{D}|} : y^*_d \leqslant y^*_{d+1}$ for all $d = 1, \ldots, |\mathcal{D}| - 1\}$ ensures that $d \mapsto Y^*_d$ is increasing and $\mathcal{S}^{IC}_{Y^*} = \mathcal{S}^I_{Y^*} \cap \{y^* \in$

[10] In this setting, Russell (2021) compared three approaches: (i) all Artstein's inequalities, (ii) the smallest available CDC based on Luo and Wang (2018), and (iii) the dual approach of Galichon and Henry (2011). Since the results of Luo and Wang (2018) did not allow intersecting conditional Artstein's inequalities over the values of the instrument, the author concluded that the CDC approach is never preferable. However, as we argued above, intersecting such inequalities is valid, so (ii) is always simpler than (i). When the smallest CDC is very large and $\mathcal{Z}$ is small, the dual approach of Galichon and Henry (2011) may be preferable. When $\mathcal{Z}$ is rich, the CDC approach is typically simpler. See Section 4.3 for a related discussion.

$\mathcal{Y}^{|\mathcal{D}|} : y^*_{d+1} - y^*_d \geqslant y^*_{d+2} - y^*_{d+1}$ for all $d = 1, \dots, |\mathcal{D}| - 2\}$ further imposes that $d \mapsto Y^*_d$ is concave. These assumptions substantially restrict the outcome space and the corresponding bipartite graphs, leading to much smaller CDCs. Panels (b) and (c) of Table 2 illustrate.

Finally, consider restricting the relationship between D and Z. Suppose that each unit in the population is characterized by a vector $D^* = (D^*_z)_{z\in\mathcal{Z}}$ of potential treatments, the observed treatment is $D = \sum_{z\in\mathcal{Z}} \mathbf{1}(Z = z) D^*_z$, and the instrument Z is jointly independent of (Y^*, D^*). Let $\mathcal{S} \subseteq \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|}$ summarize the restrictions on the outcome and treatment response functions. Given (Y, D, Z), the model produces a set-valued prediction for (Y^*, D^*)

$$G(Y, D, Z) = \{B_D(Y) \times B_Z(D)\} \cap \mathcal{S},$$

where $B_d(Y) = (\mathcal{Y} \times \dots \times \{Y\} \times \dots \mathcal{Y})$ with $\{Y\}$ in the d-th component, and $B_z(D) = (\mathcal{D} \times \dots \times \{D\} \times \dots \mathcal{D})$ with $\{D\}$ in the z-th component. Conditional on $Z = z$, the random set $G(Y, D, z)$ takes $|\mathcal{Y}||\mathcal{D}|$ distinct values, and the corresponding realizations do not have any elements in common. Thus, the corresponding bipartite graph breaks down into $|\mathcal{Y}||\mathcal{D}|$ disjoint parts corresponding to implicit-equality sets of the form $G(y, d, z)$. Then, Artstein's inequalities reduce to equalities of the form $P(Y^*_d = y, D^*_z = d) = P(Y = y, D = d \mid Z = z)$, for all $(y, d) \in \mathcal{S}$, $z \in \mathcal{Z}$, as in Balke and Pearl (1997) and Bai et al. (2024). ■

4 Implementation and Relation to Other Methods

4.1 The Master Algorithm

Algorithm 1 below summarizes the steps necessary to characterize the sharp identified set Θ_0 as in Equation (3). Throughout, we assume that X is discrete or has been discretized before defining the correspondence $G(U, X; \theta)$. We remark on continuous X below.

Algorithm 1 (Sharp Identified Set)**.**

1. **Partition the parameter space**. Fix $x \in \mathcal{X}$. Partition the parameter space, $\Theta = \bigcup_{m=1}^{M} \Theta_m(x)$, so that the support of $G(U, x; \theta)$, conditional on $X = x$, does not change with θ within each $\Theta_m(x)$. The partition can typically be constructed analytically; for linear specifications, the partition can also be obtained numerically using Algorithm 3 in Gu et al. (2022). (**Note:** this step is not always required, as discussed in detail in Section 3.4.)

2. **Partition the latent variable space**. Fix $m \in \{1, \dots, M\}$ and any $\theta \in \Theta_m$. Let $\mathcal{Y} = \{y_1, \dots, y_S\}$ denote the outcome space and $\mathcal{S}(x; \theta) = \{G_1, \dots, G_K\}$ denote the

support of $G(U, x; \theta)$, conditional on $X = x$. Partition the latent variable space as $\mathcal{U}(x, \theta) = \{u_1, \dots u_K\}$, where $u_k = \{u \in \mathcal{U} : G(u, x; \theta) = G_k\}$, and define a measure $P_{(x,\theta)}$ on $\mathcal{U}_{(x,\theta)}$ by $P_{(x,\theta)}(u_k) = P(U \in u_k \,|\, X = x)$ for all $k = 1, \dots, K$. The probabilities $P_{(x,\theta)}$ can be computed by resampling or numerical integration.

3. **Construct the bipartite graph.** Define vertices $v_1, \dots, v_S$ corresponding to $\mathcal{Y}$ and $v_{S+1}, \dots, v_{S+K}$ corresponding to $\mathcal{U}(x; \theta)$. Define the edges (v_{S+k}, v_l) for all $v_l \in G_k$, for all $k = 1, \dots, K$. Define the graph $\mathbf{B}$.

4. **Compute the smallest CDC.** Apply Algorithm 3 below to compute the smallest CDC, denoted $\mathcal{C}_m(x)$, for given m and x.

5. **Compute the identified set.** Repeating Steps 2–4, compute the classes $\mathcal{C}_m(x)$ for all $x \in \mathcal{X}$ and $m = 1, \dots, M$ to obtain Θ_0. (**Note:** In view of Corollary 1.1, for all x, θ such that the support $G(U, x; \theta)$, conditional on $X = x$, stays fixed, the graph $\mathbf{B}$, and the smallest CDC, $\mathcal{C}_m(x)$, only need to be computed once.)

The above algorithm produces a system of conditional moment inequalities of the form $\mathbb{E}[\mathbf{1}(Y \in A) - \mathbf{1}(G(U; X; \theta) \subseteq A) \,|\, X = x] \geqslant 0$, for all $A \in \mathcal{C}_m(x)$. If X is discrete or have been discretized before defining $G(U, X; \theta)$, the inequalities can be stacked together and tested using a variety of existing methods, such as Andrews and Soares (2010); Romano et al. (2014), or Cox and Shi (2023). If X is continuous, the smallest CDC approach is only practical if the support of $G(U, X; \theta)$, conditional on $X = x$, does not vary on x, so partitioning Θ in Step 1 is not required. In such settings, the resulting system of conditional inequalities can be tested using, e.g, Chernozhukov et al. (2013); Armstrong (2015), or Andrews and Shi (2017).

4.2 Computing the Smallest Core-Determining Class

Recall from Theorem 1 that the smallest CDC consists of the critical and implicit-equality sets. The latter can be found by decomposing the graph $\mathbf{B}$ into connected components, so the main challenge is to find the critical sets within each connected component. Recall that a set $A \subseteq \mathcal{Y}$ is self-connected if the subgraph of $\mathbf{B}$ induced by $(A, G^-(A))$ is connected, and complement-connected if the subgraph of $\mathbf{B}$ induced by $(A^c, G^{-1}(A^c))$ is connected. Let $N(A) = G^{-1}(A) \backslash G^-(A)$ and note that $\mathbf{B}$ is connected if and only if $N(A) \neq \varnothing$, for all A.

Say that a critical set C is a *minimal critical superset* of A if there is no critical set $\tilde{C}$ such that $A \subset \tilde{C} \subset C$. In Algorithm 2 below, we construct a correspondence $F : 2^{\mathcal{Y}} \rightrightarrows 2^{\mathcal{Y}}$ that takes a self-connected set A and returns *all* of its minimal critical supersets. By definition, such correspondence will satisfy $A \subseteq C$ for each $C \in F(A)$, and $F(\mathcal{Y}) = \varnothing$. For a collection of sets $\mathcal{C}$, define $F(\mathcal{C}) = \cup_{A \in \mathcal{C}} F(A)$. Then, in Algorithm 3, we iterate on F starting from the class $\mathcal{C} = \{G(u) : u \in \mathcal{U}\}$ until there are no more nontrivial critical supersets. Since at each step, the algorithm finds *all minimal* critical supersets, it will eventually list all critical sets.

The correspondence F is constructed as follows.

Algorithm 2 (Minimal Critical Supersets)**.**

Input: A connected bipartite graph $\mathbf{B}$ and a self-connected set A.
Output: The set of all minimal critical supersets of A.

1. Initialize $Q = \{A \cup G(u) : u \in N(A)\}$.

2. For each $C \in Q$:

   - Decompose the subgraph of $\mathbf{B}$ induced by $(C^c, G^{-1}(C^c))$ into connected components, and denote their sets of vertices by $(V_{\mathcal{Y}_l}, V_{\mathcal{U}_l})$, for $l = 1, \ldots, L$.

   - Collect all sets $C \cup \bigcup_{j \neq l} V_{\mathcal{Y}_j}$, for $l = 1, \ldots, L$, into a class $\mathcal{P}(C)$.

3. Return $\bigcup_{C \in Q} \mathcal{P}(C)$. ■

This construction is motivated by two observations. First, since any critical superset must be self-connected, it suffices to consider the sets in Q. Second, if for some $C \in Q$ the subgraph of $\mathbf{B}$ induced by $(C^c, G^{-1}(C^c))$ breaks down into several disconnected components, any minimal critical superset must contain all but one of the $V_{\mathcal{Y}_l}$ parts of these components because no other configurations can be complement-connected.

The smallest CDC is computed as follows.

Algorithm 3 (The Smallest Core-Determining Class)**.**
Input: A bipartite graph $\mathbf{B}$.
Output: The smallest core-determining class.

1. Decompose $\mathbf{B}$ into connected components $\mathbf{B}_k = ((\mathcal{Y}_k, \mathcal{U}_k), \mathcal{E}_k)$, for $k = 1, \ldots, K$.

2. For $k = 1, \ldots, K$:

   (i) Initialize $\mathcal{C}_k = \{G(u) : u \in \mathcal{U}_k\}$ and $\mathcal{R}_k = \varnothing$.

   (ii) For each $C \in \mathcal{C}_k$: check whether C is complement-connected. If so, add C to $\mathcal{R}_k$.

   (iii) Let F denote the correspondence defined by Algorithm 2. Iterate on $F(\cdot)$ starting from $\mathcal{C}_k$ and collect all sets along the way into $\mathcal{R}_k$.

3. Return $\bigcup_{k=1}^{K} \mathcal{R}_k \backslash \mathcal{Y}$. ■

We show the validity of Algorithms 2 and 3 and discuss their computational complexity in Appendix B. A major benefit of Algorithm 3 is that it is output-sensitive: its complexity is proportional to the size of the smallest CDC, as opposed to the total number of Artstein's inequalities. In general, the size of the smallest CDC may be exponential in $|\mathcal{Y}|$ and $|\mathcal{U}|$, but in many examples, it scales polynomially. The computational cost of Algorithm 3 will scale accordingly, even if the total number of Artstein inequalities becomes prohibitively large. In contrast, existing algorithms for inequality selection are based on checking each of the Artstein's inequalities for redundancy which quickly becomes computationally infeasible (see Appendix B.3 for details). Algorithm 3 can be efficiently implemented in any programming language that has a native implementation of sets (e.g., Python or Julia). For example, with our Julia implementation and MacBook Pro with M1 chip, 10 cores, and 32GB of RAM, in all examples considered in Section 3.4 in which the CDC has cardinality below 1,000, computation takes only a few seconds even in large graphs. See Appendix C.1 for details.

4.3 Comparison with Other Approaches

Besides Artstein's inequalities, several alternative approaches exist for characterizing sharp identified sets in models with set-valued predictions. Here, we describe these approaches in more detail and compare them in terms of computational tractability, obtaining sharp bounds on counterfactual quantities, and inference. Recall that $\mathcal{P}(x;\theta)$ denotes the set of model-implied distributions of the outcome Y, given $X = x$ and a parameter value $\theta \in \Theta$. Let $\mathcal{U} = \mathcal{U}(x;\theta)$ denote the partition of latent variable space given $X = x$ and θ, defined in Section 3.1. Denote $P_{Y|X=x} = (P(Y = y \,|\, X = x))_{y\in\mathcal{Y}} \in [0,1]^{|\mathcal{Y}|}$ and $P_{(x;\theta)} = (P(U \in u \,|\, X = x))_{u\in\mathcal{U}} \in [0,1]^{|\mathcal{U}|}$. To simplify exposition, we assume that X has finite support.

4.3.1 Artstein's Inequalities via Core-Determining Classes

With Artstein's inequalities, the set $\mathcal{P}(x;\theta)$ is represented as the core of the random set $G(U, x;\theta)$, conditional on $X = x$. The core is a convex compact polytope, and the smallest CDC identifies its facets. When tractable, the Artstein's inequalities approach provides a convenient characterization of the sharp identified set and has several attractive features.

First, as illustrated in Section 3.4, additional restrictions on the model — such as instrument exogeneity or outcome support restrictions — can easily be accommodated.

Second, it is theoretically straightforward to derive sharp bounds for any feature of θ_0 or a counterfactual quantity, expressed as $\phi(\theta_0)$ for some function $\phi : \Theta \to \mathbb{R}$ that is known or point-identified from the data. If Θ_0 is a connected set and ϕ is continuous, the sharp bounds on $\phi(\theta_0)$ are given by $[\min_{t\in\Theta_0}\phi(t), \max_{t\in\Theta_0}\phi(t)]$, where Θ_0 is described by a collection of moment inequalities. These optimization problems may be hard to solve in general, but when Θ_0 or ϕ have a special structure, the bounds are often easy to compute. For instance, in Example 3 above, the parameter θ represents the joint distribution of potential outcomes, so the Artstein's inequalities are linear in θ, and Θ_0 is a convex polytope. Therefore, as discussed by Russell (2021), sharp bounds on many interesting functionals of θ can be expressed via simple linear or convex optimization problems. Another class of counterfactuals for which sharp bounds are easy to compute, considered by Torgovitsky (2019) and Gu et al. (2022), is discussed in the next section.

Third, given a collection of Arstein's inequalities, inference on θ_0 or its subvectors is well-studied (see Canay and Shaikh, 2017, for a review). A minor complication arises when the CDC, and thus the set of moment inequalities to be tested, changes with θ. In such settings, as discussed in Section 3, the parameter space can be partitioned into a finite number of disjoint parts $\Theta = \bigcup_{m=1}^{M}\Theta_m$, according to the support of $G(U, X; \theta)$. Let $\hat{\phi}_{m,n}(\theta)$ be a test for $H_{0,m} : \theta \in \Theta_0(P) \cap \Theta_m$, satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_m} \sup_{\theta\in\Theta_0(P)\cap\Theta_m} \mathbb{E}_P[\hat{\phi}_{n,m}(\theta)] \leqslant \alpha,$$

for some set of distributions $\mathbf{P}_m$.[11] Then, the test $\hat{\phi}_n(\theta) = \sum_{m=1}^{M}\hat{\phi}_{m,n}(\theta)\mathbf{1}(\theta \in \Theta_m)$ for $H_0 : \theta \in \Theta_0(P)$ satisfies

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} \mathbb{E}_P[\hat{\phi}_n(\theta)] \leqslant \max_{m\in\{1,\dots,M\}} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_m} \sup_{\theta\in\Theta_0(P)\cap\Theta_m} \mathbb{E}_P[\hat{\phi}_{n,m}(\theta)] \leqslant \alpha,$$

where $\mathbf{P} = \bigcap_{m=1}^{M}\mathbf{P}_m$. As usual, the confidence set may be obtained by test inversion. Existing procedures for subvector inference (e.g., Romano and Shaikh, 2008; Bugni et al., 2017; Kaido et al., 2019) can also be modified to accommodate situations in which the set of relevant moment inequalities depends on θ. Pursuing such modifications formally is left for future research.

The test $\hat{\phi}_n(\theta)$ described above has another notable feature: it takes into account the fact that the set of moment equalities, corresponding to implicit-equality sets, may change with θ. Knowing which moment inequalities are binding is useful for inference: When constructing the test statistic, one can penalize violations in both directions, which generally leads to more powerful tests. From this perspective, $\hat{\phi}_n(\theta)$ may be preferred to a test that simply uses all Artstein's inequalities without specifying which of them are binding.

[11] The set $\mathbf{P}_m$ is typically characterized by requiring that self-normalized moment functions corresponding to the m-th part of the partition are uniformly integrable over $P \in \mathbf{P}_m$ and $\theta \in \Theta_m(P)$.

Finally, we remark that the CDC approach identifies and excludes Artstein's inequalities redundant in the population.[12] A separate question, which arises more broadly in moment inequality models, is whether the redundant inequalities can be used to improve inference procedures in finite samples. Local asymptotic analysis suggests that the answer depends on where the researcher wants to direct the power.[13] Developing a finite-sample criterion for whether to use the redundant inequalities for inference is beyond the scope of this paper, and it is an interesting direction for future research.

4.3.2 Aumann Expectation via Support Function

Beresteanu et al. (2011) represent $\mathcal{P}(x;\theta)$ as a conditional Aumann expectation of a suitable random set $Q(U,x;\theta) \subseteq \mathcal{Y}^*$, given $X = x$. Letting Y^* denote a generic integrable selection of $Q(U,x;\theta)$, the Aumann expectation $\mathbf{E}[Q(U,x;\theta)\,|\,X = x]$ is defined as the closure of the set of conditional expectations of all of its integrable selections. If the underlying probability space is non-atomic, Aumann expectation is a convex set, so it can be characterized via the support function, $h_{\mathbf{E}[Q|X=x]}(s) = \sup_{a\in\mathbf{E}[Q|X=x]} a^T s$, defined on the unit sphere $s \in S \subseteq \mathbb{R}^{|\mathcal{Y}^*|}$. The support function satisfies $h_{\mathbf{E}[Q|X=x]}(s) = \mathbb{E}[h_Q(s)|X = x]$, for all $s \in S$.[14] If the latter is easy to compute, the sharp identified set can be tractably characterized by solving, for each θ and x, a concave maximization problem in $\mathbb{R}^{d_{\mathcal{Y}^*}}$ as

$$\Theta_0 = \{\theta \in \Theta : \sup_{t\in B}(t^T\mathbb{E}[Y^*\,|\,X = x] - \mathbb{E}[h_{Q(U,x;\theta)}(t)\,|\,X = x]) \leqslant 0,\ x \in \mathcal{X} \text{ a.s.}\}. \tag{6}$$

Beresteanu et al. (2011) apply the above characterization to models with interval-valued outcomes and covariates and finite games with solution concepts other than PSNE. In such settings, using Artstein's inequalities generally does not lead to a tractable characterization of the sharp identified set.

The Aumann expectation approach can be applied in the models studied above by setting $y^*(Y) = (\mathbf{1}\{Y = y\})_{y\in\mathcal{Y}}$ and $Q(U,X;\theta) = \{y^*(Y) : Y \in G(U,X;\theta)\}$. For checking whether a given parameter value θ belongs to the sharp identified set, it often remains computationally tractable even when the smallest CDC is prohibitively large, and thus provides a viable alternative. However, other aspects of the analysis become less straightforward. First, since restricting the family of selections of $Q(U,X;\theta)$ may break the convexity of the Aumann expectation, some of the additional restrictions on the model cannot be easily accommodated; See Section 5 in Beresteanu et al. (2012) for a related discussion. Second, Equation

[12] Such inequalities are also redundant for plug-in estimation of the identified set Θ_0 or bounds on any functional $\phi(\theta_0)$. See Theorem 5.22 in Molchanov and Molinari (2018) for a related discussion.

[13] See, e.g., Example 4.1. in Canay and Shaikh (2017).

[14] See Theorems 3.4, 3.7, and 3.11 in Molchanov and Molinari (2018).

(6) describes the sharp identified set with an infinite number of conditional moment inequalities, for each $X = x$. This complicates derivations of the sharp bounds on counterfactual quantities, as well as inference procedures (see, e.g., Andrews and Shi, 2017).

4.3.3 Mixed Matching via Linear Programs or Optimal Transport

Galichon and Henry (2011) and Russell (2021) represent $\mathcal{P}(x;\theta)$ as the set of marginal distributions $P_{Y\mid X=x}$ on $\mathcal{Y}$ of all possible mixed matchings between $\mathcal{U}$ and $\mathcal{Y}$. A mixed matching is a distribution $\pi(u,y,x;\theta)$ supported on $\mathrm{Gr}(G) = \{(u,y) \in \mathcal{U} \times \mathcal{Y} : u \in G(u)\}$ that satisfies

$$\begin{aligned} \textstyle\sum_{u\in G^{-1}(y)} \pi(y,u;x,\theta) &= P_{Y|X=x}(y) \quad \text{for all } y \in \mathcal{Y}, \\ \textstyle\sum_{y\in G(u)} \pi(y,u;x,\theta) &= P_{(x;\theta)}(u) \quad \text{for all } u \in \mathcal{U}. \end{aligned} \tag{7}$$

By Farkas' Lemma, the existence of such $\pi \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{U}|}$ is equivalent to

$$\min_{\eta\in\mathbb{R}^{|\mathcal{Y}|+|\mathcal{U}|}} \big(b(x;\theta)^T\eta \,|\, A(x;\theta)^T\eta \geqslant 0\big) \geqslant 0, \tag{8}$$

where $A(x;\theta) \in \{0,1\}^{|\mathcal{Y}|\times|\mathcal{U}|} \times \{0,1\}^{|Y|+|\mathcal{U}|}$ and $b(x;\theta) \in [0,1]^{|\mathcal{Y}|+|\mathcal{U}|}$ encode the constraints in (7) and $\pi(u,y,x;\theta) \geqslant 0$ for all $(u,y) \in \mathrm{Gr}(G)$ and $\sum_{(u,y)\in\mathrm{Gr}(G)} \pi(u,y,x;\theta) = 1$. So, the sharp identified set for θ can be characterized as

$$\Theta_0 = \{\theta \in \Theta : (8) \text{ holds } x \in \mathcal{X}\text{-a.s.}\}. \tag{9}$$

Galichon and Henry (2011) propose an alternative optimal transport formulation of the problem: The goal is to transport $P_{(x,y)}(u)$ units of good from sources $u \in \mathcal{U}$ to $P_{Y|X=x}(y)$ units at terminals $y \in \mathcal{Y}$ at the minimum cost; the transportation cost is zero if $y \in G(u)$ and one otherwise. The joint distribution $\pi(u,y;x,\theta)$ satisfying (7) exists if and only if such optimal transport problem has a zero-cost solution. Modern algorithms for solving this problem have worst-case complexity of order $(|\mathcal{Y}| + |\mathcal{U}|) \times |E(\mathbf{B})|$; see, e.g., Orlin (2013).[15]

The mixed matching approach sometimes remains computationally tractable when the smallest CDC is not, and thus provides another viable alternative. Additional modeling assumptions can be accommodated, although less conveniently than with the CDC approach. For example, consider imposing independence of the latent variables $U \in \mathcal{U}$ and an excluded

[15] As another alternative, Galichon and Henry (2011) propose using submodular minimization. The sharp identified set for θ can be expressed as $\Theta_0 = \{\theta \in \Theta : \min_{A\subseteq\mathcal{Y}} F_{(x;\theta)}(A) \geqslant 0, x \in \mathcal{X}\text{-a.s.}\}$, where $F_{(x;\theta)} = P(Y \in A \,|\, X = x) - C_{G(U;x,\theta)}(A)$. Since $F_{(x;\theta)}(\cdot)$ is submodular, the above minimization problem is often feasible. For each x, ignoring the cost of evaluating $C_{G(U,x;\theta)}(A)$, the worst-case complexity of the above problem is $|\mathcal{Y}|^6$; see, e.g., Orlin (2009). This method appears to be generally slower than the optimal transport approach, unless $|\mathcal{U}| \gg |\mathcal{Y}|^3$.

instrument $Z \in \mathcal{Z}$, as in Example 3 discussed in Section 3.4.[16] With the CDC approach, conditional Artstein's inequalities can simply be intersected over Z. With the mixed matching approach, to ensure that the $\mathcal{U}$-marginal of π is independent of Z, additional $|\mathcal{Z}|-1$ matching constraints are required for each $u \in \mathcal{U}$. When $|\mathcal{Z}|$ is large or infinite, the task becomes infeasible. In terms of bounding counterfactual quantities, the mixed matching approach is applicable if the parameter of interest can be expressed directly in terms of π. In the context of Example 3, Russell (2021) provides evidence the linear programs describing sharp bounds on certain functionals of the joint distribution of potential outcomes scale favorably with $|\mathcal{Y}|$ for fixed $|\mathcal{D}|$ and $|\mathcal{Z}|$. More generally, similar to the support function approach, Equations (8)–(9) describe the identified set by an infinite number of conditional moment inequalities, which complicates derivations of the sharp bounds on counterfactual quantities, as well as inference procedures.

4.3.4 Minimal Relevant Partition

A closely related approach for characterizing sharp bounds on a class of counterfactuals in discrete-outcome models using linear programming was proposed by Tebaldi et al. (2019) and Gu et al. (2022). In Gu et al. (2022), the model consists of the factual outcome and random set, $Y \in G(U, X; \theta)$, and the counterfactual outcome and random set $Y^* \in G^*(U, X; \theta)$. The parameter of interest is a linear functional of the counterfactual distribution of Y^*, conditional on X, denoted $\phi(P_{Y^*|X})$. The counterfactual set of predictions G^* is assumed to be "coarser" than the factual set G in the following sense: There must exist a finite partition $\{u_1^*, \ldots, u_L^*\}$ of the latent variable space $\mathcal{U}$ such that knowing the probabilities of "cells" u_l^*, conditional on $X = x$, suffices to bound $\phi(P_{Y^*|X})$. Following Tebaldi et al. (2019), such a partition is called the Minimal Relevant Partition (MRP). Similarly to the mixed matching approach, the authors show that $Y \in G(U, X; \theta)$, a.s., and $Y^* \in G^*(U, X; \theta)$, a.s., hold jointly (with all random quantities defined on a common probability space) if and only if there exists a joint mixed matching $\pi_x(y, y^*, u_l^*)$ consistent with the model. That is, $\pi_x(y, y^*, u_l^*)$ is the probability that a factual outcome y is chosen from the set $G(u_l^*, x; \theta)$, a counterfactual outcome y^* is chosen from the set $G^*(u_l^*, x; \theta)$, and $u \in u_l^*$, conditional on $X = x$. Such a structure enables the authors to express sharp bounds on the counterfactual $\phi(P_{Y^*|X^*})$ via two linear programs. The choice vector in these programs, $(\pi_x(y, y^*, u_l^*))_{y, y^* \in \mathcal{Y}, x \in \mathcal{X}, l \leqslant L}$, is of dimension $d = |\mathcal{X}||\mathcal{Y}|^2 L$, and there are $p = |\mathcal{X}|(|\mathcal{Y}| + 2)$ constraints to ensure that $\pi_x(y, y^*, u_l^*)$ matches the observed conditional distribution of the outcomes and represents a valid probability distribution and $q = |\mathcal{X}||\mathcal{Y}|^2 L$ non-negativity constraints.[17]

[16] To match the notation in this section and Example 3, let $U = Y^*$, $X = \varnothing$, and $Y = (Y, D)$.

[17] See Section 2.2 in Gu et al. (2022)

The CDC approach can also be applied in this framework, and it sometimes leads to simpler linear programs. The idea is to treat the probabilities of "cells" in the MRP, denoted $\mu(u_l^*, x)$, as unknown parameters. Such "cells" are typically finer than the partition $\mathcal{U}(x;\theta) = \{u_1, \ldots, u_k\}$ described in Section 3.1, so each $\mu(u_k, x)$ is a sum of several $\mu(u_l^*, x)$. Artstein's inequalities provide linear inequality constraints on $\mu(u_k, x)$ of the form $P(Y \in A \mid X = x) \geqslant \sum_{k \in G^-(A)} \mu(u_k, x)$, for all $A \in \mathcal{C}^*(x)$. Assuming, for example, that $\mathcal{C}^*(x)$ does not change with x, this approach leads to a linear program with the choice vector $(\mu(u_l^*, x))_{x \in \mathcal{X}, l \leqslant L}$ of dimension $d = |\mathcal{X}|L$, $p = |\mathcal{X}|K$ equality constraints linking the MRP with $\mathcal{U}(x;\theta)$, and $q = |\mathcal{X}|(|\mathcal{C}^*(x)| + L)$ inequality constraints including the Artstein's inequalities and non-negativity constraints. Then, if $|\mathcal{C}^*(x)|$ is smaller than $|\mathcal{Y}|^2$, the resulting linear program is easier than the one described in the preceding paragraph. In particular, this is the case in many entry games in Example 1 and a dynamic entry model in Example 2.

4.3.5 Final Remarks

To summarize the above discussion, when the smallest CDC is manageable, Artstein's inequalities approach provides a simple and universally applicable method for deriving sharp identified sets for both structural parameters and counterfactuals. It is especially useful in settings with excluded exogenous covariates that have rich support and are independent of the unobservables. When the smallest CDC is very large, other methods discussed above provide viable alternatives.

5 Extensions: Infinite Support, Dominated Selections

In this section, we extend the results of Section 3 to models in which the outcome variable has infinite support, possibly with some additional restrictions. Such settings require a more nuanced formal setup, which we now intoduce.

Let Q be a sigma-finite measure on $\mathcal{Y}$ and suppose that in addition to $Y \in G(U, X; \theta)$, almost surely, the researcher wishes to impose that the distribution of Y is absolutely continuous with respect to Q. For example, choosing to Q to be a Lebesgue measure imposes that Y has a continuous distribution, and restricting the support of Q corresponds to restricting the support of Y. As in Section 3.1, we shall fix $x \in \mathcal{X}$ and $\theta \in \Theta$ and work with the random set $G(U, x; \theta)$, conditional on $X = x$. Notice $P_Y \ll Q$ implies $P_{Y|X=x} \ll Q$, for almost all $x \in \mathcal{X}$. For simplicity, we denote $G \equiv G(U, x; \theta)$, $P \equiv P_{U \mid X=x,\theta}$, and $C_G(A) = P(G(U, x; \theta) \subseteq A \mid X = x)$. We can then view G as a closed random set defined on the probability space $(\mathcal{U}, \mathcal{F}, P)$, where $\mathcal{F}$ is the Borel sigma-field on $\mathcal{U}$, and taking values in $(\mathcal{Y}, \mathcal{B})$. Recall that $\mathfrak{C}$ denotes the class of all closed subsets of $\mathcal{Y}$ and let $\mathcal{M}_Q$ denote the set of all probability distributions on $(\mathcal{Y}, \mathcal{B})$ absolutely continuous with respect to Q.

For any class of sets $\mathcal{C} \subseteq \mathfrak{C}$, define

$$\mathcal{M}_Q(\mathcal{C}) = \{\mu \in \mathcal{M}_Q : \mu(A) \leqslant C_G(A), \text{ for all } A \in \mathcal{C}\}.$$

Our object of interest will be the set of distributions of all selections of G that are absolutely continuous with respect to Q, or, in the above notation,

$$\mathcal{M}_Q(\mathfrak{C}) = \text{Core}(G) \cap \mathcal{M}_Q.$$

We assume that Q is chosen to ensure that $\mathcal{M}_Q(\mathfrak{C}) \neq \varnothing$, i.e, for any $B \in \mathfrak{C}$ with $Q(B) = 0$, $C_G(B) = 0$. By analogy with Definition 2.1, we introduce the following notion.

Definition 5.1 (Q-CDC)**.** *A class $\mathcal{C} \subseteq \mathfrak{C}$ is Q-core-determining if $\mathcal{M}_Q(\mathcal{C}) = \mathcal{M}_Q(\mathfrak{C})$.*

Generally, $\mathcal{M}_Q(\mathfrak{C}) \subseteq \text{Core}(G)$, although in many settings, the two sets are equal. The following example illustrates.

Example 4 (Dominated Selections)**.** Let $\mathcal{U} = [0,1]^2$ and $\mathcal{Y} = [0,1]$ both be endowed with Borel sigma-fields. Let $U = (U_1, U_2)$ be a pair of random variables with a joint distribution P supported on $S \subseteq S_0 = \{(u_1, u_2) \in [0,1]^2 : u_1 \leqslant u_2\}$. Consider a random closed set $G : [0,1]^2 \rightrightarrows [0,1]$ defined by $G(U) = [U_1, U_2]$.

Depending on P, the random set G may have only continuous selections or a full menu including continuous, discrete, and mixed selections. For example, if P is any continuous distribution will full support S_0, the random set G can be arbitrarily narrow with positive probability, so it only has continuous selections. That is, $\text{Core}(G) = \mathcal{M}_\lambda(\mathfrak{C})$ with λ being the Lebesgue measure. Alternatively, suppose $U_2 = U_1 + 1/K$, P-almost-surely, for some $K \in \mathbb{N}$. Then, for example, $Y = U_1$ is a continuous selection of G, and $Y' = \sum_{k=0}^{K-1} \frac{k+1}{K} \mathbf{1}(U_1 \in \left[\frac{k}{K}, \frac{k+1}{K}\right))$ is a discrete selection of G. In this case, taking Q to be the Lebesgue measure on [0,1] will meaningfully restrict the set of selections. ■

Given the measures Q and P, each set $A \in \mathcal{B}$ can be associated with an equivalence class $[A]$ with $A' \sim A$ if $A = A'$, Q-a.s., and $G^-(A) = G^-(A')$, P-a.s.. For the purpose of describing the core, all sets $A \in [A]$ are equivalent. Therefore, we define the critical and implicit equality sets as follows.

Definition 5.2 (Critical and Implicit-Equality Sets)**.** *A set $A \in \mathfrak{C}$ is critical if $\mathcal{M}_Q(\mathfrak{C} \backslash [A]) \neq \mathcal{M}_Q(\mathfrak{C})$. A set $A \in \mathfrak{C} \backslash \{\mathcal{Y}, \varnothing\}$ is an implicit equality set if $\mu(A) = C_G(A)$ for all $\mu \in \mathcal{M}_Q(\mathfrak{C})$.*

Since the containment functional is uniquely defined by the family of closed sets $\mathfrak{C}$, it is natural to think of a representative closed set $A \in \mathfrak{C}$ in each equivalence class $[A]$. In what follows, we write A instead of $[A]$ and speak of sets instead of equivalence classes, for simplicity. For any sets $A, B \in \mathcal{B}$, we say that "$A = B$, Q-a.s.," if $Q((A \cap B^c) \cup (B \cap A^c)) = 0$. For any sets $A, B \in \mathcal{F}$, define "$A = B$, P-a.s" similarly.

The results in Section 3 relied on connectivity of the bipartite graph $\mathbf{B}$ or its subgraphs. A direct analog of the bipartite graph $\mathbf{B}$ in the present setting is the *graph* of G:

$$\mathrm{Gr}(G) = \{(u, y) \in \mathcal{U} \times \mathcal{Y} : y \in G(u)\}.$$

However, requiring that this set be connected is not sufficient for our purposes, because the connections have to be "detectable" by the measures Q and P. Recall that the lower and upper pre-images of G are defined as

$$G^-(A) = \{u \in \mathcal{U} : G(u) \subseteq A\}; \qquad G^{-1}(A) = \{u \in \mathcal{U} : G(u) \cap A \neq \varnothing\}$$

and $G^-(A) \subseteq G^{-1}(A)$, for each $A \subseteq \mathcal{Y}$. Further, let

$$N(A) = G^{-1}(A) \backslash G^-(A) = \{u \in \mathcal{U} : G(u) \cap A \neq \varnothing, G(u) \cap A^c \neq \varnothing\}$$

be the set vertices $u \in \mathcal{U}$ that connect A with the rest of the graph $\mathrm{Gr}(G)$. We define connected graphs as follows.

Definition 5.3 (Connected Graph of a Random Set)**.** *A random set G has a connected graph if (i) $Q(G(u)) > 0$, for P-almost all $u \in \mathcal{U}$; For any $A \in \mathfrak{C} \backslash \mathcal{Y}$ with $Q(A) > 0$: (ii) $P(N(A)) > 0$, and (iii) For almost all $u \in N(A)$, $Q(G(u) \cap A) > 0$ and $Q(G(u) \cap A^c) > 0$.*

In the finite setting studied in Section 3, connectivity amounts to $N(A) \neq \varnothing$, for all A, while $G(u) \neq \varnothing$, for all $u \in \mathcal{U}$, and $G(u) \cap A \neq \varnothing$ and $G(u) \cap A^c \neq \varnothing$, for all $u \in N(A)$, hold by definition. Assumptions (i)–(iii) above additionally require that the respective sets are "detectable" by the measure Q. If $P(N(A)) = 0$, for some A, the outcome space can be partitioned as $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$, with $\mathcal{Y}_1 = A$ and $\mathcal{Y}_2 = A^c$, so that $G^{-1}(\mathcal{Y}_1) \cap G^{-1}(\mathcal{Y}_2) = \varnothing$, P-a.s.. That is, the correspondence G "breaks" into two P-a.s. disjoint components, which can be analyzed separately. In complete models, G is singleton-valued, so $Q(G(u)) = 0$ is possible even if $G(u) \neq \varnothing$, and $G^{-1}(A) = G^-(A)$ and $N(A) = \varnothing$, for all $A \subseteq \mathcal{Y}$, so the set $\mathrm{Gr}(G)$ breaks into a potentially infinite number of disjoint pieces. The following example illustrates.

Example 4 (Continued. Graph-Connected Random Sets)**.** Assume the same setup as above. Let Q be the Lebesgue measure. Let P be any continuous distribution supported on S_0. For

P-almost all u, the set $G(u)$ has positive length, and so $Q(G(u)) > 0$. Consider, for example, a set $A = [a_1, a_2]$ for some $0 < a_1 < a_2 < 1$. Then, $N(A) = \{(u_1, u_2) \in S : u_1 \in A, u_2 > a_2\} \cup \{(u_1, u_2) \in S : u_1 < a_1, u_2 \in A\}$. For P-almost all $u \in N(A)$, the segments $G(u) \cap A$ and $G(u) \cap A^c$ have positive length and thus $Q(G(u) \cap A) > 0$ and $Q(G(u) \cap A^c) > 0$. Therefore, the graph of G is connected in the sense of Definition 5.3. Alternatively, suppose P is supported on the union of sets $S_1 = \{(u_1, u_2) \in S_0 : u_2 \leqslant 1/2\}$ and $S_2 = \{(u_1, u_2) \in S_0 : u_1 \geqslant 1/2\}$. Then, $G^{-1}([0, 1/2]) = S_1$ and $G^{-1}([1/2, 1]) = S_2$. Then, the graph of random set G is not connected in the sense of Definition 5.3. In this case, the restrictions $G_1 : S_1 \to [0, 1/2]$ and $G_2 : S_2 \to [1/2, 1]$ can be considered separately. ■

Finally, the notions of self- and complement-connected sets extend as follows.

Definition 5.4 (Self- and Complement-Connected Sets)**.** *Let G be a random set with a connected graph, in the sense of Definition 5.3. A subset $A \in \mathfrak{C}$ is self-connected if there do not exist A_1, A_2 satisfying $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \varnothing$, Q-a.s., and $G^-(A) = G^-(A_1) \cup G^-(A_2)$, P-a.s.. A subset $A \in \mathfrak{C}$ is complement-connected if there do not exist A_1, A_2 satisfying $A^c = A_1^c \cup A_2^c$ and $A_1^c \cap A_2^c = \varnothing$, Q-a.s., and $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \varnothing$, P-a.s..*

5.1 The Smallest Core-Determining Class

We are now ready to state the main results of this section, which are direct extensions of Lemmas 1 and 2 and Theorem 1.

Lemma 3 (Critical Sets)**.** *Let $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set with a connected graph. A subset $A \in \mathfrak{C}$ is critical if and only if it is self- and complement-connected.*

Lemma 4 (Implicit-Equality Sets)**.** *Let $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set. Suppose there is a countable partition $\mathcal{Y} = \bigcup_{l \geqslant 1} \mathcal{Y}_l$ such that $\mathcal{Y}_i \cap \mathcal{Y}_j = \varnothing$, Q-a.s., and $G^{-1}(\mathcal{Y}_i) \cap G^{-1}(\mathcal{Y}_j) = \varnothing$, P-a.s., for all $i \neq j$, and such partition cannot be further refined. A subset $A \subseteq \mathcal{Y}$ is an implicit-equality set if and only if $A = \bigcup_{l \in L_A} \mathcal{Y}_l$ for some $L_A \subseteq \mathbb{N}$.*

Theorem 2 (Smallest CDC)**.** *Let $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set.*

1. *If G has connected graph, the class $\mathcal{C}^*$ of all critical sets, characterized in Lemma 3, is the smallest CDC.*

2. *If the outcome space $\mathcal{Y}$ can be partitioned as in Lemma 4, and $\mathcal{C}_l^*$ denotes the class of all critical sets in $\mathcal{Y}_l$, characterized in Lemma 3, then $\mathcal{C}^* = \bigcup_{l \geqslant 1} \mathcal{C}_l^* \cup \mathcal{Y}_l$ is the smallest CDC, up to removing a single arbitrary $\mathcal{Y}_l$.*

Corollary 1.1 and the subsequent discussion also apply in continuous-outcome settings. When the support of $G(U, x; \theta)$, conditional on $X = x$, is infinite, the smallest CDC, $\mathcal{C}^*(x, \theta)$ contains an infinite number of sets, for each x. This fact implies that using all of the modeling implication for estimation and inference on θ_0 may be challenging.[18] On the other hand, certain functionals of the form $\phi(\theta_0) \in \mathbb{R}$, may admit relatively simple sharp bounds. In such cases, Theorem 2 can be used to "guess" the sharp bounds, but to prove sharpness, it is typically easier to explicitly construct a data-generating distribution that attains the bounds. The following examples illustrate.

5.2 Examples

The first example studies a model with interval-valued data. For related results, see Beresteanu et al. (2012), Section 2.3 in Molinari (2020), and Manski (1994).

Example 5 (Interval Data)**.** Let $Y^* \in \mathcal{Y}$ denote a continuous outcome variable (e.g., income) and $X \in \mathcal{X}$ denote covariates (e.g., socio-economic characteristics). Suppose the researcher does not observe Y^* directly but observes continuous random variables $Y_L, Y_U \in \mathcal{Y}$ such that $Y^* \in G(Y_L, Y_U) = [Y_L, Y_U]$ (e.g., income bracket). For simplicity, suppose X is discrete, and $\mathcal{Y} = [\underline{y}, \overline{y}]$ for some known $\underline{y} < \overline{y}$. Also, suppose that $P(\underline{\kappa}(x) \leqslant Y_U - Y_L \leqslant \overline{\kappa}(x) \,|\, X = x) = 1$ for some known functions $\underline{\kappa}(x)$ and $\overline{\kappa}(x)$. The basic parameter of interest is $\theta_0 = P_{Y^* X}$.

Consider the random set $G(Y_L, Y_U)$, conditional on $X = x$. Since Y^* is continuous, we take Q equal to the Lebesgue measure on $\mathcal{Y}$. Since the conditions of Definition 5.3 are satisfied, the random set G is graph-connected, and there are no implicit-equality sets. In turn, the critical sets can be determined as follows. The support of G is the set of all closed intervals in $[\underline{y}, \overline{y}]$. The only sets that satisfy $A = G(G^-(A))$, i.e., can be expressed as unions of elements of the support of G, are finite or countable unions of disjoint intervals included in $[\underline{y}, \overline{y}]$, where each interval has a length of at least $\underline{\kappa}(x)$. Consider a union of the form $A = A_1 \cup A_2 = [a_1, b_1] \cup [a_2, b_2]$ with $b_j - a_j \geqslant \underline{\kappa}(x)$ and $a_2 > b_1$. Then, $A_1 \cap A_2 = \varnothing$ and $G^-(A) = G^-(A_1) \cup G^-(A_2)$, P-a.s, meaning that A is not self-connected. A similar argument applies to any other collection of disjoint intervals, which means that all critical sets must be contiguous intervals. Next, consider an interval $A = [a, b]$ with $\underline{y} < a < b < \overline{y}$ and $b - a > \overline{\kappa}(x)$. Then, the sets $A_1 = [\underline{y}, b]$ and $A_2 = [a, \overline{y}]$ satisfy $A_1^c \cup A_2^c = A^c$, $A_1^c \cap A_2^c = \varnothing$, and $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \varnothing$, P-a.s., meaning that A is not complement-connected. Note that intervals of the form $[\underline{y}, b]$ and $[a, \overline{y}]$ are complement-connected. Thus, the sharp identified set for θ_0 is completely characterized by inequalities of the form $P(Y^* \in A \,|\, X = x) \geqslant$

[18] However, e.g., Mourifié and Wan (2017) show that the local average treatment effect assumptions in a model with continuous outcomes can be tested using the procedure of Chernozhukov et al. (2013).

$P([Y_L, Y_U] \subseteq A \,|\, X = x)$ for all sets A in the class

$$\mathcal{C}^*(x) = \{[\underline{y}, a], [a, \overline{y}] : \underline{y} + \underline{\kappa}(x) \leqslant a \leqslant \overline{y} - \underline{\kappa}(x)\} \cup \{[a, b] : \underline{\kappa}(x) \leqslant b - a \leqslant \overline{\kappa}(x)\},$$

for all $x \in \mathcal{X}$. If $\underline{\kappa}$ or $\overline{\kappa}$ do not depend on x or its subvector, the corresponding inequalities can be intersected. Importantly, Theorem 2 implies that each of the above inequalities is also necessary to guarantee sharpness.

Next, suppose the parameter of interest is the conditional CDF $\phi(\theta_0) = F_{Y^*|X=x}(\cdot)$. The sharp identified set for $\phi(\theta_0)$ is contained in the "tube" of non-decreasing functions satisfying

$$F_{Y^*|X=x}(y) \in \begin{cases} [0, F_{Y_L|X=x}(\underline{\kappa}(x))] & y \in [0, \underline{\kappa}(x)) \\ [F_{Y_U|X=x}(y), F_{Y_L|X=x}(y)] & y \in [\underline{y} + \underline{\kappa}(x), \overline{y} - \underline{\kappa}(x)] \\ [F_{Y_U|X=x}(\overline{y} - \underline{\kappa}(x)), 1] & y \in (\overline{y} - \kappa(x), \overline{y}]. \end{cases}$$

The upper and lower bounds correspond to valid CDF's and are sharp. However, not all CDFs inside the tube are included in the sharp identified set, because valid candidates must also satisfy

$$F_{Y^*|X=x}(b) - F_{Y^*|X=x}(a) \geqslant P(Y_L \geqslant a, Y_U \leqslant b \,|\, X = x) \tag{10}$$

for any a, b such that $\underline{\kappa}(x) \leqslant b - a \leqslant \overline{\kappa}(x)$. This rules out CDFs that increase "too little" over any such interval. Importantly, Theorem 2 implies that no other restrictions are required.

Finally, suppose the parameter of interest is the difference between conditional quantiles $\phi(\theta_0) = q_{Y^*|X=x}(\tau_1) - q_{Y^*|X=x}(\tau_2)$, for some $\tau_1 > \tau_2$. Each of the quantiles is sharply bounded by the corresponding quantiles of Y_L and Y_U, which may suggest that

$$\phi(\theta_0) \in \left[\max\{0, q_{Y_L|X=x}(\tau_1) - q_{Y_U|X=x}(\tau_2)\},\ q_{Y_U|X=x}(\tau_1) - q_{Y_L|X=x}(\tau_2)\right].$$

However, the upper bound may not be sharp due to (10) being violated at $a = q_{Y^*|X=x}(\tau_2)$, $b = q_{Y^*|X=x}(\tau_1)$. Instead, it can be verified that the sharp upper bound is

$$\max\{b - a \,|a \geqslant q_{Y_L|X=x}(\tau_2), b \leqslant q_{Y_U|X=x}(\tau_1),\ \tau_1 - \tau_2 \geqslant P(Y_L \geqslant a, Y_U \leqslant b|X = x)\}.$$

Bounds on other functionals can be obtained similarly. ■

Our final example is a model of ascending auctions studied by Haile and Tamer (2003), Aradillas-López et al. (2013), Chesher and Rosen (2017), and Molinari (2020).

Example 6 (Ascending Auctions)**.** Consider an ascending auction with N bidders. Let $V_j \in [0, \overline{v}]$ and $B_j \in [0, \overline{v}]$ denote the valuation and bid of player j, and $V_{j:N}$ and $B_{j:N}$ denote the corresponding j-th smallest valuation and bid. Suppose the bidders are symmetric in the sense that $(V_1, \ldots, V_N)$ are exchangeable. Let $F \in \mathcal{F}$ denote the joint distribution of ordered valuations $V = (V_{1:N}, \ldots, V_{N:N})$ supported on $S = \{v \in [0, \bar{v}]^N : v_1 \leqslant \cdots \leqslant v_N\}$, where the class $\mathcal{F}$ summarizes the assumptions on the information structure. Suppose there is no reserve price and minimal bid increment. Suppose the researcher observes the two largest bids $(B_{N-1:N}, B_{N:N})$ and wants to learn about features of F. Following Haile and Tamer (2003), suppose that bidders (i) do not bid above their valuation and (ii) do not let their opponents win at a price they would be willing to pay. Then, (i) implies $B_{j:N} \leqslant V_{j:N}$ for all j, and (ii) implies $V_{N-1:N} \leqslant B_{N:N}$. Thus, the model produces a set-valued prediction for the bids, given valuations, $G(V; F) = [0, V_{N-1:N}] \times [V_{N-1:N}, V_N] \cap S$. As long as F is supported on S, the support of $G(V; F)$ does not depend on F.

It can be verified that the random set $G(V; F)$ has connected graph, so by Lemma 4, there are no implicit-equality sets. In turn, the class of all critical sets is vast. In particular, it includes all lower sets $A_1 = \{(v_1, v_2) \in [0, \bar{v}]^2 : v_1 \leqslant \kappa(v_2)\}$, for some weakly decreasing function $\kappa : [0, \bar{v}] \to [0, \bar{v}]$; all sets of the form $A_2 = \{(v_1, v_2) : v_1 \leqslant a, v_2 \in [b, c]\}$, for some $a, b, c \in [0, \bar{v}]$ with $b \leqslant c$; all sets of the form $A_1 \cap A_2$; and all countable unions of the resulting family of sets. As a result, the sharp identified set for F is intractable in practice. However, certain functionals of F admit tractable bounds. Aradillas-López et al. (2013) show that in ascending auctions, if the transaction price equals the largest of the reserve price and second-highest valuation, the expected profit and bidders' surplus under counterfactual reserve prices depend only on the marginal distribution of the two largest valuations: $\phi(F) = (F_{N-1:N}, F_{N:N})$. The sharp identified set for $\phi(F)$ is given by

$$\Phi_0 = \{\phi(F) : F \in \mathcal{F}, P((B_{N-1:N}, B_{N:N}) \in A) \geqslant P_F([0, V_{N-1:N}] \times [V_{N-1:N}, V_N] \subseteq A)\ \forall A\}.$$

To make progress, Aradillas-López et al. (2013) assume that the valuations are positively dependent in the sense that the probability $P(V_i \leqslant v \,|\, \#\{j \neq i : V_j \leqslant v\} = k)$ is non-decreasing in k for each $i = 1, \ldots, N$. Under the above assumption, the authors show that $F_{N:N} \in [F_{N-1:N}, \phi_{N-1:N}(F_{N-1:N})^N]$, where $\phi_{N-1:N} : [0, 1] \to [0, 1]$ is a known strictly increasing function that maps the distribution of the second-largest order statistic of an i.i.d. sample of size N to the parent distribution. With this assumption, the set Φ_0 can be characterized more concretely. The Artstein's inequality corresponding to the set $A = S \cap [0, v] \times [0, \bar{v}]$ implies $F_{N-1:N}(v) \leqslant G_{N-1:N}(v)$; the set $A = S \cap [0, \bar{v}] \times [v, \bar{v}]$ implies $F_{N-1:N}(v) \geqslant G_{N:N}(v)$; and the set $A = S \cap [0, v] \times [0, v]$ implies $F_{N:N}(v) \leqslant G_{N:N}(v)$.

Combining these inequalities with the bounds on $F_{N:N}$ yields

$$\begin{aligned} G_{N:N}(v) \leqslant F_{N-1:N}(v) &\leqslant G_{N-1:N}(v); \\ \phi_{N-1:N}(G_{N:N}(v))^N \leqslant \quad F_{N:N}(v) &\leqslant G_{N:N}(v). \end{aligned}$$

By constructing suitable joint distributions $F \in \mathcal{F}$, it is possible to show that both upper bounds and both lower bounds can be attained simultaneously, so the bounds are sharp.

As in the preceding example, although the bounds on $F_{N-1:N}$ are sharp, the corresponding "tube" of functions includes many CDFs that do not belong to the sharp identified set. Specifically, the set $A = S \cap [a, \overline{v}] \times [0, b]$ for $b > a$ corresponds to the Artstein's inequality $F_{N-1:N}(b) - F_{N-1:N}(a) \geqslant P(B_{N-1:N} \geqslant a, B_{N:N} \leqslant b)$, which rules out CDFs that do not increase sufficiently between a and b. This fact has immediate implications for studying, e.g., optimal reserve prices. The details are left for future research. ■

6 The Importance of Selecting Inequalities

In this section, we provide evidence that selecting Artstein's inequalities informally may lead to a substantial loss of identifying information.

Dynamic Entry In the first simulation exercise, we revisit the dynamic entry model of Berry and Compiani (2020), which is our Example 2. In this setting, even with only a few time periods, the total number of Artstein's inequalities is prohibitively large; see Table 1b. To this end, the authors suggest using inequalities that should intuitively be informative about the structural parameters. Specifically, they consider the events: "the firm enters at least once," "the firm exits at least once," and "the number of firms in the market does not change for K consecutive periods." Below, we compare the resulting identified sets with the sharp identified set for $T = 5$ made feasible by computing the smallest CDC.

The true parameter values are set to $\bar{\pi} = 0.5$, $\gamma = 1.5$, and $\rho = 0.75$, and the sample size is 10,000. Further details of the simulation design are provided in Appendix C. Figure 5 presents the results. The grey shaded regions represent projections of the sharp identified set in the model with $T = 2$; the orange regions combine the inequalities for $T = 2$ with the hand-picked inequalities of Berry and Compiani (2020) for $T = 5$; and the light-blue regions correspond to the sharp identified set with $T = 5$. Evidently, the intuitive inequalities do not come close to using all of the identifying information in the model with $T = 5$. In numerical terms, the orange ("intuitive") identified set for (π, γ, ρ) is roughly 26% smaller than the grey one, while the blue (sharp) identified set is 97% smaller.

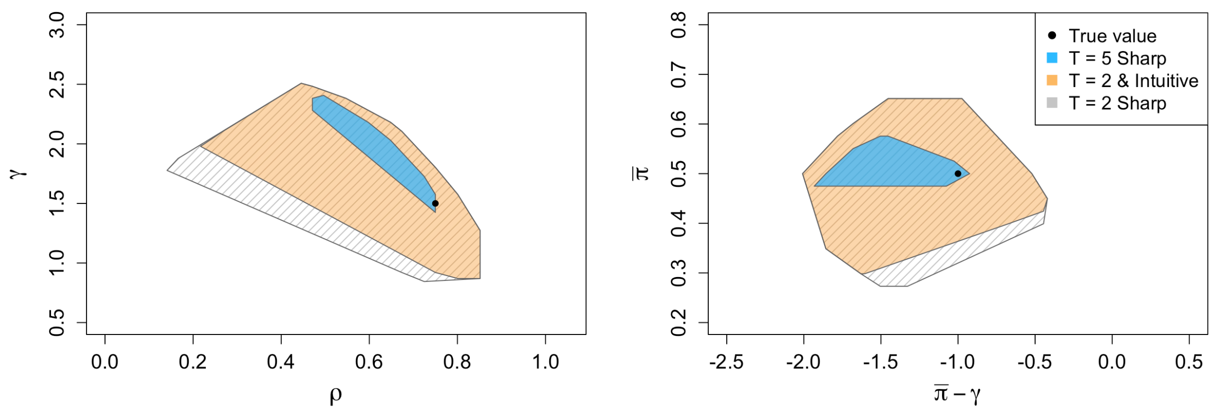


Figure 5: Projections of identified sets in the dynamic entry model from Example 2.

Static Entry In the second simulation exercise, we aim to quantify how much identifying information would be lost if the researcher used alternative sets of inequalities instead of the smallest CDC. We revisit the market entry model from Example 1 with $N = 3$ players and strategic complementarities, $\delta_j > 0$ for $j \in \{1, 2, 3\}$. In this setting, there are 254 nontrivial Artstein's inequalities in total, while the smallest CDC contains only 14 inequalities. A comprehensive experiment would require trying all sets of 14 inequalities out of 254 ($\approx 10^{22}$ options), which is computationally infeasible. As an approximation, we sample 14 out of 254 inequalities at random 15,000 times and compute the corresponding identified sets using a fixed grid of points. For each set of inequalities, we compute the relative size of the sharp identified set to the simulated one as the ratio of the counts of grid points that satisfy the respective inequalities. We simulate 5,000 observations with parameters $\alpha_j = -0.4$ and $\delta_j = 0.4$ and unobservables ε_j distributed i.i.d. $N(0, 1)$, for $j \in \{1, 2, 3\}$. Within the regions of multiplicity, we select asymmetric equilibria (e.g., $(1, 1, 0)$ instead of $(0, 0, 0)$) with probability 0.9 to ensure that each outcome is realized with a non-trivial probability. The resulting distribution over $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1), (0, 0, 0), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ is $(0.08, 0.08, 0.08, 0.25, 0.25, 0.09, 0.09, 0.08)$. The grid for (α, δ) is $[-0.5, -0.2] \times [0.3, 0.5]$ with 50 values along each dimension.

Figure 6 presents the results. The left panel depicts the sharp identified set, and the right panel shows the distribution of the relative size of the sharp identified set across simulations. The median relative size of the sharp identified set to the simulated ones is 38%, meaning that in half of the simulations at least 62% of the identifying information is lost. This result suggests that the smallest CDC is a very specific collection of inequalities and using alternative sets of inequalities is likely to result in a substantial loss of identifying information.

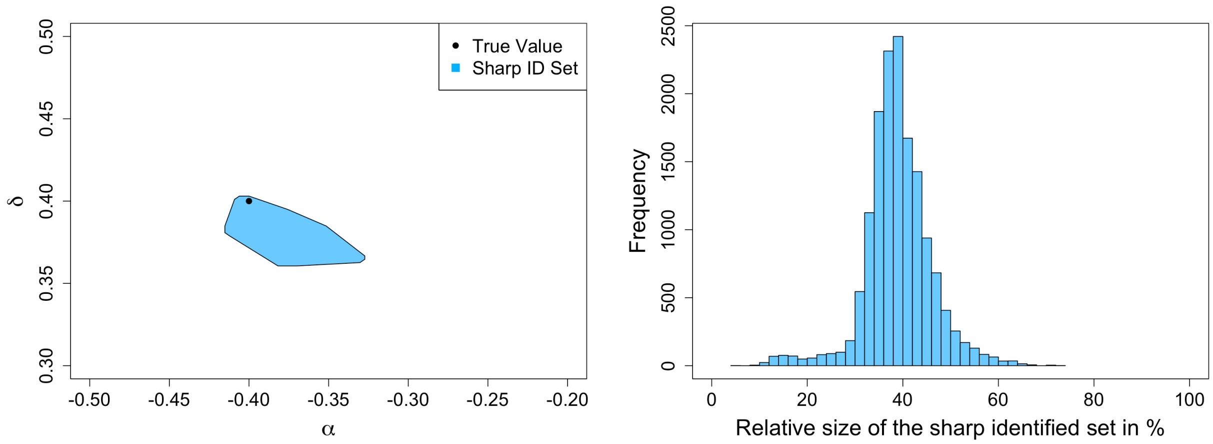


Figure 6: Size of the sharp identified set relative to identified sets constructed with the same number of inequalities in a market entry model with complementarities in Example 1.

7 Conclusion

Artstein's inequalities provide a convenient way to describe sharp identified sets in a large class of partially-identified econometric models. However, the total number of inequalities is often prohibitively large in practice, while many of them are redundant in the sense that excluding them from the analysis is without loss of identifying information. In this paper, we derived the smallest possible set of inequalities that suffices for sharpness, provided an efficient algorithm to compute it, and used the proposed approach to obtain tractable characterizations of the sharp identified sets in several well-studied settings. The results apply far beyond the examples considered in the paper. Determining which moment inequalities are more informative for inference in finite samples is an important question for future research.

References

Andrews, D. W. and Shi, X. (2017). Inference based on many conditional moment inequalities. *Journal of Econometrics*, 196(2):275–287.

Andrews, D. W. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157.

Aradillas-López, A. (2020). The econometrics of static games. *Annual Review of Economics*, 12(1):135–165.

Aradillas-López, A., Gandhi, A., and Quint, D. (2013). Identification and inference in ascending auctions with correlated private values. *Econometrica*, 81(2):489–534.

Armstrong, T. B. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics*, 186(1):51–65.

Artstein, Z. (1983). Distributions of random sets and random selections. *Israel Journal of Mathematics*, 46(4):313–324.

Avis, D. and Fukuda, K. (1991). A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. In *Proceedings of the seventh annual symposium on Computational geometry*, pages 98–104.

Bai, Y., Huang, S., Moon, S., Shaikh, A. M., and Vytlacil, E. J. (2024). On the identifying power of monotonicity for average treatment effects.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American statistical Association*, 92(439):1171–1176.

Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048.

Beresteanu, A., Molchanov, I., and Molinari, F. (2008). Sharp identification regions in games. Technical report, cemmap working paper.

Beresteanu, A., Molchanov, I., and Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6):1785–1821.

Beresteanu, A., Molchanov, I., and Molinari, F. (2012). Partial identification using random set theory. *Journal of Econometrics*, 166(1):17–32.

Berry, S. and Tamer, E. (2006). Identification in models of oligopoly entry. *Econometric Society Monographs*, 42:46.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, pages 889–917.

Berry, S. T. and Compiani, G. (2020). An instrumental variable approach to dynamic models. Technical report, National Bureau of Economic Research.

Bontemps, C. and Kumar, R. (2020). A geometric approach to inference in set-identified entry games. *Journal of Econometrics*, 218(2):373–389.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of political economy*, 99(5):977–1009.

Bugni, F. A., Canay, I. A., and Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, 8(1):1–38.

Canay, I. A. and Shaikh, A. M. (2017). Practical and theoretical advances in inference for partially identified models. *Advances in Economics and Econometrics*, 2:271–306.

Chernozhukov, V., Lee, S., and Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, 81(2):667–737.

Chesher, A. and Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, 85(3):959–989.

Chesher, A. and Rosen, A. M. (2020). Generalized instrumental variable models, methods, and applications. In *Handbook of Econometrics*, volume 7, pages 1–110. Elsevier.

Chesher, A., Rosen, A. M., and Smolinski, K. (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4(2):157–196.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Cox, G. and Shi, X. (2023). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *The Review of Economic Studies*, 90(1):201–228.

De Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263–288.

Galichon, A. and Henry, M. (2011). Set identification in models with multiple equilibria. *The Review of Economic Studies*, 78(4):1264–1298.

Gu, J., Russell, T., and Stringham, T. (2022). Counterfactual identification and latent space enumeration in discrete outcome models. *Available at SSRN 4188109*.

Gualdani, C. (2021). An econometric model of network formation with an application to board interlocks between firms. *Journal of Econometrics*, 224(2):345–370.

Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1):1–51.

Heckman, J. J., Smith, J., and Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, 6:4779–4874.

Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, 87(4):1397–1432.

Kédagni, D., Li, L., and Mourifié, I. (2020). Discordant relaxations of misspecified models. *arXiv preprint arXiv:2012.11679*.

Kleinberg, J. and Tardos, E. (2006). *Algorithm Design*. Addison Wesley.

Kline, B., Pakes, A., and Tamer, E. (2021). Moment inequalities and partial identification in industrial organization. In *Handbook of Industrial Organization*, volume 4, pages 345–431. Elsevier.

Luo, Y. and Wang, H. (2018). Identifying and computing the exact core-determining class. *Available at SSRN 3154285*.

Manski, C. F. (1994). The selection problem. In *Advances in Econometrics, Sixth World Congress*, volume 1, pages 143–70.

Manski, C. F. (2003). *Partial identification of probability distributions*, volume 5. Springer.

Manski, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review*, 48(4):1393–1410.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997–1010.

Manski, C. F. and Pepper, J. V. (2009). More on monotone instrumental variables. *The Econometrics Journal*, 12:S200–S216.

Mbakop, E. (2023). Identification in some discrete choice models: A computational approach. *arXiv preprint arXiv:2305.15691*.

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. *Journal of econometrics*, 195(2):224–235.

Molchanov, I. and Molinari, F. (2018). *Random sets in econometrics*, volume 60. Cambridge University Press.

Molchanov, I. S. (2005). *Theory of random sets*, volume 19. Springer.

Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of econometrics*, 7:355–486.

Mourifié, I. and Wan, Y. (2017). Testing local average treatment effect assumptions. *Review of Economics and Statistics*, 99(2):305–313.

Orlin, J. B. (2009). A faster strongly polynomial time algorithm for submodular function minimization. *Mathematical Programming*, 118(2):237–251.

Orlin, J. B. (2013). Max flows in o (nm) time, or better. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 765–774.

Pakes, A., Porter, J., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1):315–334.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, 138(9):2786–2807.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002.

Russell, T. M. (2021). Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, 39(2):532–546.

Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Tebaldi, P., Torgovitsky, A., and Yang, H. (2019). Nonparametric estimates of demand in the california health insurance exchange. Technical report, National Bureau of Economic Research.

Telgen, J. (1983). Identifying redundant constraints and implicit equalities in systems of linear constraints. *Management Science*, 29(10):1209–1222.

Torgovitsky, A. (2019). Partial identification by extending subdistributions. *Quantitative Economics*, 10(1):105–144.

Tsukiyama, S., Ide, M., Ariyoshi, H., and Shirakawa, I. (1977). A new algorithm for generating all the maximal independent sets. *SIAM Journal on Computing*, 6(3):505–517.

A Proofs from the Main Text

Proof of Lemma 1 The "Only if" direction follows from the arguments in Section 3.2: if a set A is not self-connected, the first argument applies; if A is not comlement-connected, the second argument applies. For the "If" direction, let $\nu \in \mathcal{M}$ with $\nu(y) > 0$ for all $y \in \mathcal{Y}$. Let A be a set that is both self-connected and complement connected. Define a map $\pi_A : \mathcal{U} \times 2^{\mathcal{Y}} \to [0,1]$ via

$$\pi_A(u; B) = \frac{\nu(G(u) \cap B \cap A^c)}{\nu(G(u) \cap A^c)} \mathbf{1}(u \in N(A)) + \frac{\nu(G(u) \cap B)}{\nu(G(u))} \mathbf{1}(u \notin N(A))$$

Note that $u \in N(A)$ ensures $G(u) \cap A^c \neq \varnothing$. By standard properties of measurable functions, the map $u \mapsto \pi_A(u; B)$ is measurable, for each B. By construction, for each fixed u, $\pi_A(u; \cdot)$ is a probability distribution on $\mathcal{Y}$ supported on $G(u)$. That is, π_A is a Markov kernel. Note that $\pi_A(u, B) > 0$ if and only if $u \in G^-(B) \cup N(B)$, and $\pi_A(u, B) = 1$ for $u \in G^-(B)$. Averaging over u yields a probability distribution $\mu_A(B) = \sum_{u \in \mathcal{U}} \pi_A(u; B) P(u)$ satisfying

$$\mu_A(B) = C_G(B) + \sum_{u \in N(B) \cap N(A)} \frac{\nu(G(u) \cap B \cap A^c)}{\nu(G(u) \cap A^c)} P(u) + \sum_{u \in N(B) \cap N(A)^c} \frac{\nu(G(u) \cap B)}{\nu(G(u))} P(u). \tag{A.1}$$

In particular, $\mu_A(A) = C_G(A)$. We will show that for any $B \neq A$, the second or the third summand (or both) in (A.1) must be positive, so $\mu_A(B) > C_G(B)$.

Since G is connected, it must be that $N(B) \neq \varnothing$, for all B. If $N(B) \cap N(A)^c \neq \varnothing$, the last summand in (A.1) is strictly positive and the conclusion follows. It remains to consider $N(B) \subseteq N(A)$. There are three possible cases:

1. $A \cap B \neq \varnothing$ and $A \cap B^c \neq \varnothing$. Since $N(B) \subseteq N(A)$, in particular, $G^-(A) \cap N(B) = \varnothing$. That is, all u such that $G(u) \subseteq A$ satisfy either $G(u) \cap B = \varnothing$ (i.e., $G(u) \subseteq B^c$) or $G(u) \subseteq B$. Thus, the sets $A_1 = A \cap B$, $A_2 = A \cap B^c$ satisfy $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \varnothing$, and $G^-(A_1) \cup G^-(A_2) = G^-(A)$, which contradicts the assumed self-connectivity of A.

2. $A \cap B = \varnothing$, or $A \subseteq B^c$. Then, $B \cap A^c = B$, so the second summand in (A.1) is positive.

3. $A \cap B^c = \varnothing$, or $A \subseteq B$. Since $N(B) \subseteq N(A)$, B is connected to the rest of the graph only through A. That is, there does not exist u such that $G(u) \cap (B \cap A^c) \neq \varnothing$ and $G(u) \cap B^c \neq \varnothing$. So, for all u such that $G(u) \cap (B \cap A^c) \neq \varnothing$ it must be that $G(u) \cap B^c = \varnothing$, and for all u such that $G(u) \cap B^c \neq \varnothing$, it must be that $G(u) \cap (B \cap A^c) = \varnothing$. Thus, the sets $A_1^c = B \cap A^c$, $A_2^c = B^c \cap A^c = B^c$ satisfy $A^c = A_1^c \cup A_2^c$ and $G^{-1}(A_1^c) \cup G^{-1}(A_2^c) = G^{-1}(A^c)$, which contradicts the complement-connectivity of A.

Therefore, we have constructed a probability measure $\mu_A \in \text{Core}(G)$ satisfying $\mu_A(A) = C_G(A)$ and $\mu_A(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$. By continuity, there exists a probability measure μ' such that $\mu'(A) = C_G(A) - \epsilon$ for some small $\epsilon > 0$, while $\mu'(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A, \tilde{A} \in \mathfrak{C}$. Such μ' satisfies $\mu' \notin \mathcal{M}(\mathfrak{C})$, and $\mu' \in \mathcal{M}(\mathfrak{C} \backslash A)$. Therefore, A must be critical. ■

Proof of Lemma 2 For the "If" direction, let Y be an arbitrary selection of G with a distribution μ. Since for each $l \in \{1, \ldots, L\}$, $Y \in \mathcal{Y}_l$ holds if and only if $U \in G^-(\mathcal{Y}_l)$, it must be that $\mu(\mathcal{Y}_l) = P(U \in G^-(\mathcal{Y}_l)) = C_G(\mathcal{Y}_l)$. By additivity of probability measures, the corresponding equality holds for any union of sets $\mathcal{Y}_l$.

For the "Only if" direction, let A be any set other than a union of some $\mathcal{Y}_l$. By assumption, each subgraph $\mathbf{B}_l$ induced by $(\mathcal{Y}_l, G^{-1}(\mathcal{Y}_l))$ is connected, so it must be that $N(A) \neq \varnothing$. Let $\nu \in \mathcal{M}$ with $\nu(y) > 0$ for all $y \in \mathcal{Y}$ and define a Markov kernel $\pi_0 : \mathcal{U} \times 2^{\mathcal{Y}} \to [0, 1]$ as $\pi_0(u; A) = \nu(A \cap G(u))/\nu(G(u))$. For each $u \in \mathcal{U}$, $\pi_0(u; \cdot)$ is a probability measure supported on $G(u)$ satisfying $\pi_0(u; A) > 0$ if and only if $u \in G^-(A) \cup N(A)$, with $\pi_0(u; A) = 1$, for $u \in G^-(A)$. Averaging over u yields a probability distribution

$$\mu_0(A) = \sum_{u \in \mathcal{U}} \pi_0(u; A) P(u) = C_G(A) + \sum_{u \in N(A)} \pi_0(u; A) P(u).$$

Since $N(A) \neq \varnothing$, $\pi_0(u; A) > 0$, and $P(u) > 0$ for all $u \in N(A)$, it follows that $\mu_0(A) > C_G(A)$, so such A cannot be an implicit-equality set. ■

Proof of Theorem 1 Let $\subseteq$ and $\subset$ denote the weak and strict inclusions correspondingly. First, suppose $\mathbf{B}$ is connected. Since any core-determining class must contain all critical sets, the goal is to show that the class of all critical sets is itself core-determining. To this end, it suffices to show that removing any non-critical set A cannot "make" any other non-critical set A' critical, i.e., $\mathcal{M}(\mathfrak{C} \backslash A) = \mathcal{M}(\mathfrak{C} \backslash A') = \mathcal{M}(\mathfrak{C})$ necessarily implies $\mathcal{M}(\mathfrak{C} \backslash A \backslash A') = \mathcal{M}(\mathfrak{C})$.

For each set of vertices $S \subseteq V(\mathbf{B})$, let $\mathcal{N}(S)$ denote the set of all edges adjacent to some vertex in S. To each set $A \subseteq \mathcal{Y}$, associate a set of edges adjacent to A^c or $G^-(A)$, that is,

$$\mathcal{E}_A \equiv \mathcal{N}(A^c \cup G^-(A)) = \{(u, y) \in E(\mathbf{B}) : y \in A^c \text{ or } u \in G^-(A)\}.$$

If A is non-critical, by Lemma 1, either (i) $A = A_1 \cup A_2$ with $A_1 \cap A_2 = \varnothing$ and $G^-(A) = G^-(A_1) \cup G^-(A_2)$ or (ii) $A^c = A_1^c \cup A_2^c$ with $A_1^c \cap A_2^c = \varnothing$ and $G^{-1}(A^c) = G^{-1}(A_1^c) \cup G^{-1}(A_2^c)$. In either case, it can be verified that $\mathcal{E}_A \subset \mathcal{E}_{A_j}$ for $j = 1, 2$. For example, if (i) holds, $\mathcal{E}_A = \mathcal{N}(G^-(A_1)) \cup \mathcal{N}(G^-(A_2)) \cup \mathcal{N}(A^c)$, while $\mathcal{E}_{A_1} = \mathcal{N}(G^-(A_1)) \cup \mathcal{N}(A_2) \cup \mathcal{N}(A^c)$. By construction, $\mathcal{N}(G^-(A_2)) \subseteq \mathcal{N}(A_2)$. Since the graph $\mathbf{B}$ is connected, there must be an edge between A_2 and $\mathcal{U} \backslash G^-(A)$. That is, $\mathcal{N}(G^-(A_2)) \subset \mathcal{N}(A_2)$ and therefore $\mathcal{E}_A \subset \mathcal{E}_{A_1}$. The inclusion $\mathcal{E}_A \subset \mathcal{E}_{A_2}$ is symmetric, and case (ii) can be considered similarly. Since in either case, removing A is without loss as long as A_1 or A_2 are present, the fact that $\mathcal{E}_A \subset \mathcal{E}_{A_1}$ and $\mathcal{E}_A \subset \mathcal{E}_{A_2}$ implies that removing a non-critical set A cannot make any other set $A' \neq A$ critical. Otherwise, we would have $\mathcal{E}_A \subset \mathcal{E}_{A'}$ and $\mathcal{E}_{A'} \subset \mathcal{E}_A$, which is a contradiction.

Next, let $\mathcal{Y} = \bigcup_{l=1}^L \mathcal{Y}_l$ with $\mathcal{Y}_i \cap \mathcal{Y}_j = \varnothing$ for $i \neq j$, denote the finest partition of the outcome space with the property $G^{-1}(\mathcal{Y}_i) \cap G^{-1}(\mathcal{Y}_j) = \varnothing$. Then, any set of the form $A = \cup_{l=1}^L A_l$ with $A_l \subseteq \mathcal{Y}_l$ satisfies $G^-(A) = \bigcup_{l=1}^L G^-(A_l)$, so it is redundant given $(A_l)_{l=1}^L$ (see also Theorem 2.33 in Molchanov and Molinari, 2018). Also, since $\sum_{l=1}^L \mu(\mathcal{Y}_L) = 1$ for any $\mu \in \text{Core}(G)$, any one (and only one) of the sets $\mathcal{Y}_l$ can be omitted from the CDC. Combining these facts with the above argument applied to each connected component $\mathbf{B}_l$ of $\mathbf{B}$ yields the result. ■

Proof of Lemma 3 Let $\nu \in \mathcal{M}_Q$ with $d\nu/dQ > 0$. Let A be a set that is both self-connected and complement connected. Define a map $\pi_A : \mathcal{U} \times \mathcal{B} \to [0, 1]$ as

$$\pi_A(u; B) = \frac{\nu(G(u) \cap B \cap A^c)}{\nu(G(u) \cap A^c)} \mathbf{1}(u \in N(A)) + \frac{\nu(G(u) \cap B)}{\nu(G(u))} \mathbf{1}(u \notin N(A)).$$

Since G has a connected graph, $\nu(G(u)) > 0$, for almost all $u \in \mathcal{U}$, and $\nu(G(u) \cap A^c) > 0$, for almost all $u \in N(A)$. By the Robbins' Theorem (Theorem 1.5.16 in Molchanov, 2005) and standard properties of measurable functions, the map $u \mapsto \pi_A(u, B)$ is measurable for each $B \in \mathcal{B}$. By construction, for each fixed u, $\pi_A(u; \cdot)$ is a probability distribution on $\mathcal{Y}$ supported on $G(u)$. That is, π_A is a Markov kernel. Note that $\pi_A(u; B) > 0$ if and only if $u \in G^-(B) \cup N(B)$, and $\pi_A(u; B) = 1$, for any $u \in G^-(B)$. Averaging over u yields a probability measure $\mu_A(B) = \int_{\mathcal{U}} \pi_A(u; B) dP(u)$, satisfying

$$\mu_A(B) = C_G(B) + \int_{N(B) \cap N(A)} \frac{\nu(G(u) \cap B \cap A^c)}{\nu(G(u) \cap A^c)} dP(u) + \int_{N(B) \cap N(A)^c} \frac{\nu(G(u) \cap B)}{\nu(G(u))} dP(u). \tag{A.2}$$

In particular, $\mu_A(A) = C_G(A)$ holds by construction. We will show that, for any $B \neq A$ with

$Q(B) > 0$, at least one of the remaining summands in (A.2) must be positive, so $\mu_A(B) > C_G(B)$. The inequalities for B with $Q(B) = 0$ hold trivially provided that $\mathcal{M}_Q(\mathfrak{C}) \neq \varnothing$, so such sets need not be considered.

An integral of the form $\int_S f(\omega)dP(\omega)$, where $f(\omega) \geq 0$, is positive if and only if $P(\{\omega : f(\omega) > 0\} \cap S) > 0$. For each $B \subseteq \mathcal{Y}$, the set $N(B)$ contains "vertices" $u \in \mathcal{U}$ that connect B with the rest of the graph $\text{Gr}(G)$. Since G has connected graph, $P(N(B)) > 0$, for all B with $Q(B) > 0$. If $P(N(B) \cap N(A)^c) > 0$, the last summand in (A.1) is strictly positive and the conclusion follows. The rest of the argument proceeds exactly as in the proof of Lemma 1, with qualifiers P-a.s. and Q-a.s. added when referring to set operations in $\mathcal{U}$ and $\mathcal{Y}$. ■

Proof of Lemma 4 The proof is nearly identical to that of Lemma 2 with the following modifications. The measure $\nu \in \mathcal{M}_Q$ must satisfy $d\nu/dQ > 0$ and the measurability of $u \mapsto \pi_0(u; A)$ follows from the Robbins theorem, as in the proof of Lemma 3. The qualifiers P-a.s. and Q-a.s. are added when referring to set operations in $\mathcal{U}$ and $\mathcal{Y}$. ■

Proof of Theorem 2 The proof is nearly identical to that of Theorem 1. To each set $A \in \mathfrak{C}$, we can associate a collection of "egdes"

$$\mathcal{E}_A = \{(u, y) \in \text{Gr}(G) : y \in A^c \text{ or } u \in G^-(A)\}.$$

Suppose G is graph-connected and a set A is non-critical. Then, it must be that either $A = A_1 \cup A_2$ with $G^-(A) = G^-(A_1) \cup G^-(A_2)$ or $A^c = A_1^c \cup A_2^c$ with $G^{-1}(A^c) = G^{-1}(A_1^c) \cup G^{-1}(A_2^c)$. In either case, it can be verified that $\mathcal{E}_A \subset \mathcal{E}_{A_1}$ and $\mathcal{E}_A \subset \mathcal{E}_{A_2}$ with both inclusions being "detectable" in the sense that $Q(\{y : (u, y) \in \mathcal{E}_{A_j} \backslash \mathcal{E}_A\}) > 0$ and $P(\{u : (u, y) \in \mathcal{E}_{A_j} \backslash \mathcal{E}_A\}) > 0$, for $j \in \{1, 2\}$. This observation implies that removing a non-critical set A cannot make any other set A' critical. Indeed, assuming otherwise would imply that $\mathcal{E}_A \subset \mathcal{E}_{A'}$ and $\mathcal{E}_{A'} \subset \mathcal{E}_A$, which is a contradiction. ■

B Algorithms 2 and 3

B.1 Validity

It suffices to show that Algorithm 2 identifies all minimal critical supersets of a given self-connected set. By Lemma 1, critical sets must be self-and complement-connected. Given a self-connected set A, the idea is to list all possible expansions of A, denoted $C = A \cup B$, that satisfy two properties: (i) C is self- and complement-connected and (ii) there is no self- and complement-connected $\tilde{C}$ such that $A \subset \tilde{C} \subset C$ with strict inclusions. To be self-connected, the set C must contain $G(u)$ for some $u \in G^{-1}(A) \backslash G^-(A)$. To find a minimal such C, it suffices to look for $C = A \cup G(u)$ for $u \in G^{-1}(A) \backslash G^{-}(A)$. If the subgraph of $\mathbf{B}$ induced by $(C^c, G^{-1}(C^c))$ is connected, such C is one of the minimal critical supersets of A. If this subgraph "breaks" into disconnected components, denoted here by $(\mathcal{Y}_l, \mathcal{U}_l, \mathcal{E}_l)$, for $l = 1, \ldots, L$, then only sets of the form $P_l = C \cup \bigcup_{j \neq l} \mathcal{Y}_j$, for some l, can be minimal critical sets. Indeed, such P_l is self-connected because each of $\mathcal{Y}_j$ must be linked with C (otherwise, the graph $\mathbf{B}$ would be disconnected), and complement-connected since the subgraph induced by $(P_l, G^{-1}(P_l))$ is precisely the remaining connected component $(\mathcal{Y}_l, \mathcal{U}_l, \mathcal{E}_l)$. Also, any proper subset of P_l cannot be complement-connected by construction. Therefore, Algorithm 2 finds all minimal critical supersets. That Algorithm 3 finds all critical sets follows from the discussion in the main text.

B.2 Computational Complexity

The time complexity of decomposing the graph $\mathbf{B}$ into connected components using Depth First Search is $|V(\mathbf{B})| + |E(\mathbf{B})|$, where $|V(\mathbf{B})| = |\mathcal{Y}| + |\mathcal{U}|$ and $|E(\mathbf{B})|$ are the numbers of vertices and edges in $\mathbf{B}$ correspondingly (see, e.g., Section 3.2 in Kleinberg and Tardos, 2006). All further calculations apply within each connected component. To keep notation simple, we assume that $\mathbf{B}$ itself is connected.

Let $|\mathcal{C}^*|$ denote the size of the smallest core-determining class. Let $\overline{N} = \max_{A \subseteq \mathcal{Y}} |N(A)|$ denote the maximum cardinality of the set of vertices connecting a self-connected set $A \subseteq \mathcal{Y}$ with the rest of the graph $\mathbf{B}$, and $\overline{L}$ denote the maximum number of connected components of the subgraph of $\mathbf{B}$ induced by $(A^c, G^{-1}(A^c))$. These quantities are trivially bounded by $\overline{N} < |\mathcal{U}|$ and $\overline{L} < |\mathcal{Y}|$ but are often much smaller and may remain bounded in large graphs.

First, consider Algorithm 2. Step 1 of the Algorithm requires reading at most $\overline{N}$ sets from the adjacency list of $\mathbf{B}$. For each of these sets, Step 2 decomposes a subgraph of $\mathbf{B}$ into connected components and creates a list of at most $\overline{L}$ sets as a result. The complexity of decomposing a subgraph of $\mathbf{B}$ into connected components is bounded by that of decomposing the whole graph $\mathbf{B}$. Thus, the complexity of Step 2 is bounded by $\overline{N} \cdot (|\mathcal{Y}| + |\mathcal{U}| + |E(\mathbf{B})| + \overline{L} \cdot |\mathcal{Y}|)$. The number of minimal critical supersets of A is bounded by $\overline{L} \cdot \overline{N}$, and each of them has size at most $|\mathcal{Y}|$, so removing the duplicates in Step 3 has complexity $\overline{L} \cdot \overline{N} \cdot |\mathcal{Y}|$. Therefore, the total complexity of Algorithm 2 is bounded above by $\overline{N} \cdot (|\mathcal{Y}| + |\mathcal{U}| + |E(\mathbf{B})| + \overline{L} \cdot |\mathcal{Y}|)$.

Next, consider Algorithm 3. Step 1 has complexity $|\mathcal{Y}| + |\mathcal{U}| + |E(\mathbf{B})|$, as discussed above. Step 2-(i) requires reading $|\mathcal{U}|$ sets from the adjacency list of $\mathbf{B}$ and Step 2-(ii) requires checking complement connectivity for every such set, which has complexity $|\mathcal{U}| + |\mathcal{Y}| + |E(\mathbf{B})|$, so the the total complexity is $|\mathcal{U}| \cdot (|\mathcal{U}| + |\mathcal{Y}| + |E(\mathbf{B})|)$. Let $\mathcal{C}$ denote the output of Step 2-(i). Step 2-(iii) applies Algorithm 2 first to all sets A in $\mathcal{C}$, and then iteratively to the resulting collection of sets, denoted $\mathcal{C}'$. It may be the case that $F(A) \cap F(A') \neq \varnothing$ for two distinct sets $A, A' \in \mathcal{C}'$, so computing the union $F(\mathcal{C}') = \bigcup_{A \in \mathcal{C}'} F(A)$ requires removing duplicate sets after each iteration. Every set $F(A)$ contains at most $\overline{L} \cdot \overline{N}$ sets of size at most $|\mathcal{Y}|$, and there are at most $\max(|\mathcal{U}|, |\mathcal{C}^*|)$ elements in each $\mathcal{C}'$, so the complexity of removing the duplicates is bounded by $\overline{L} \cdot \overline{N} \cdot |\mathcal{Y}| \cdot \max(|\mathcal{U}|, |\mathcal{C}^*|)$. Since the duplicates are eliminated, at every iteration — except possibly the first — Algorithm 2 is only applied to critical sets. Taking stock, the complexity of Algorithm 3 is bounded by

$$\overline{L} \cdot \overline{N} \cdot \max(|\mathcal{U}|, |\mathcal{C}^*|) \cdot (|\mathcal{Y}| + |\mathcal{U}| + |E(\mathbf{B})|). \tag{B.1}$$

Assuming bounded $\overline{L}$ and $\overline{N}$ and $|\mathcal{U}| \leqslant |\mathcal{C}^*|$, all of which typically hold in applications, the complexity bound in (B.1) is comparable with that of an oracle algorithm which receives critical sets one by one and verifies that each of them is self- and complement- connected.

B.3 Connections with Existing Algorithms

One way to compute the smallest CDC (denoted as $\mathcal{C}^*$) is to start from the system of all Artstein's inequalities and remove redundant ones. There exist generic methods for identifying redundant and implicit-equality constraints in linear systems (see, e.g., Telgen, 1983; Schrijver, 1998). In practice, such methods require solving one linear program per constraint, so the resulting algorithmic complexity scales proportionally to the total number of constraints. In many settings, the total number of Artstein's inequalities is exponential in $|\mathcal{Y}|$ and $|\mathcal{U}|$, which quickly makes the above approach computationally infeasible. In contrast, our approach does not require considering each of the potentially redundant constraints. Instead, it uses the additional structure of the problem (i.e., the bipartite graph) to directly "build" the non-redundant constraints, thus substantially lifting the computational burden.

Another way to compute the smallest CDC is to find a minimal half-space representation of a polytope given in a vertex representation (recall Section 4.3.3). The relevant polytope is $\mathcal{P} = \{A\pi : \pi \geqslant 0, \pi'\mathbf{1} = 1\}$, where A is a $(|\mathcal{Y}| + |\mathcal{U}|) \times |E(\mathbf{B})|$ binary matrix, in which each column $a_e \in \{0,1\}^{|\mathcal{Y}|+|\mathcal{U}|}$ represents an edge $e = (u, y) \in E(\mathbf{B})$. Artstein's inequalities provide a half-space representation of $\mathcal{P}$, which may contain redundant elements, and the smallest CDC gives a minimal such representation. More generally, a minimal half-space representation of any polytope can be computed numerically using the algorithm of Avis and Fukuda (1991). For the so-called "non-degenerate" problems, in which every facet of $\mathcal{P}$ contains exactly $|\mathcal{Y}| + |\mathcal{U}|$ vertices, their algorithm is shown to be output-sensitive: the complexity of recovering the $|\mathcal{C}^*|$ facets of $\mathcal{P}$ is $O(|\mathcal{C}^*| \cdot (|\mathcal{Y}| + |\mathcal{U}|) \cdot E(\mathbf{B}))$. In our setting, the "non-degeneracy" means that for every set $A \in \mathcal{C}^*$, there are exactly $|\mathcal{Y}| + |\mathcal{U}|$ edges

$(u, y) \in E(\mathbf{B})$ such that $\mathbf{1}(y \in A^c) + \mathbf{1}(u \in G^-(A)) = 1$. This requirement appears to be very restrictive and is not satisfied in any of our examples. Although the algorithm of Avis and Fukuda (1991) applies more broadly, it is not generally output-sensitive and often does not scale well in practice. In turn, our Algorithm 3 applies only to polytopes associated with bipartite graphs and computes the minimal half-space representation in an output-sensitive manner without any further restrictions.

Algorithm 3 is also related to the problem of listing all maximal independent sets (MIS) of vertices in a graph. A MIS is a set of vertices that are mutually disconnected, and such that adding any other vertex would violate this condition. Tsukiyama et al. (1977) proposed an output-sensitive algorithm for listing all MIS-s in an arbitrary undirected graph Γ with a worst-case complexity $O(|MIS| \times |V(\Gamma)| \times |E(\Gamma)|)$. In our setting, each MIS corresponds to a set of the form $(A^c, G^-(A))$, where $A \subseteq \mathcal{Y}$ can be expressed as a union of elements of the support of G (see Section 3.2). If all such A are self- and complement-connected, the set of all MIS is the smallest CDC, and the worst-case complexity of our algorithm matches that of Tsukiyama et al. (1977). Otherwise, Algorithm 3 finds MIS satisfying further connectivity restrictions.

C Further Simulation Evidence

C.1 Computing Times

Tables 3 and 4 below summarize the computing times for all examples in the main text where computation was done numerically. All computation was performed in Julia on a 2021 MacBook Pro with M1 chip with 10 cores and 32 GB RAM.

C.2 Dynamic Entry Game

Our simulation design follows that of Berry and Compiani (2020). Let T be the number of observed periods and $\bar{T} = 50 + T$ the total number of periods used in the simulation. Let $N = 10{,}000$ be the sample size. The data are generated as follows: (i) Draw N vectors of latent variables ε of size $\bar{T}$ according to the $AR(1)$ process specified in Example 2; (ii) For each sample, draw $X_1 \sim$ Bernoulli($p = 0.5$) and solve for the optimal policy for $\bar{T}$ periods. (iii) Keep the last T periods as the observed data. There are three main parameters $(\bar{\pi}, \gamma, \rho)$ set to $(0.5, 1.5, 0.75)$, and an auxiliary parameter $\pi' = \pi - \gamma = -1$. The grid has step size 0.025 and boundaries $\pi \in [-1.5, 1.5]$, $\pi' \in [-3, 0]$, and $\rho \in [0, 1]$.

Heterogeneous firms, $\delta_j > 0$					
N	2	3	4	5	6
# Vertices	9	23	71	275	1341
# Edges	6	22	100	534	3320
Computing time (in seconds)	10^{-4}	10^{-3}	10.81	–	–
Two types of firms, $\beta_3 = \beta_2$					
(N^1, N^2)	(1, 1)	(2, 2)	(2, 4)	(2, 7)	(6, 6)
# Vertices	9	20	32	50	84
# Edges	6	15	25	40	71
Computing time (in seconds)	10^{-4}	10^{-4}	10^{-4}	10^{-3}	10^{-3}
Two types of firms, $\beta_3 > \beta_2$					
# Vertices	9	23	42	69	135
# Edges	6	21	49	88	221
Computing time (in seconds)	10^{-4}	10^{-3}	10^{-3}	10^{-3}	0.08

(a) Entry games in Example 1.

T	2	3	4	5	6	7	8	9	10
# Vertices	15	31	63	127	255	511	1023	2047	4095
# Edges	14	30	62	126	254	510	1022	2046	4094
Computing time (in seconds)	10^{-4}	10^{-3}	10^{-3}	0.02	0.09	0.42	1.86	8.61	43.16

(b) Dynamic binary choice model from Example 2.

Table 3: Graph characteristics and computing times for the smallest CDC in Examples 1–2

Notes: Computing times are averaged over 10 runs. Symbol "–" indicates that a single run did not finish within 1 hour. See tables 1a and 1b for the results.

Monotone outcome response							
$\lvert\mathcal{D}\rvert \setminus \lvert\mathcal{Y}\rvert$	2	3	4	5	6	7	8
2	10^{-4}	10^{-4}	10^{-4}	10^{-3}	0.02	0.11	0.69
3	10^{-4}	10^{-3}	0.01	0.15	2.42	33.7	505.46
4	10^{-4}	10^{-3}	0.12	4.7	167.33	–	–
Monotone and concave outcome response							
$\lvert\mathcal{D}\rvert \setminus \lvert\mathcal{Y}\rvert$	2	3	4	5	6	7	8
3	10^{-4}	10^{-4}	10^{-3}	0.03	0.37	4.76	59.94
4	10^{-4}	10^{-4}	10^{-3}	0.1	2.04	45.05	1044.93

Table 4: Computing times for the smallest CDC in the potential outcomes model from Example 3

Notes: Computing times (in seconds) are averaged across 10 runs. Symbol "–" indicates that a single run did not finish within 1 hour. See Table 2 for the results.

D Additional Examples

We have reserved two more examples for the appendix. The first example is a discrete choice model with endogenous covariates, studied by Chesher et al. (2013) and Tebaldi et al. (2019).

Example 4 (Discrete Choice with Endogeneity)**.** Individuals choose one of $J+1$ alternatives, $Y \in \{y_0, y_1, \dots, y_J\} \equiv \mathcal{Y}$, where y_0 represents the outside option. Choosing y_j yields utility $v_j(X) + \varepsilon_j$, where $X \in \{x_1, \dots, x_K\} \equiv \mathcal{X}$ may include prices and individual- and market-level covariates, and $\varepsilon_j \in \mathbb{R}$ are latent utility shifters. Individuals maximize their utility, so $Y = y_{j^*}$ for $j^* = \operatorname{argmax}_j\{v_j(X) + \varepsilon_j\}$. Normalize $v_0(x) = 0$, for all x, and $\varepsilon_0 = 0$. Some components of X may be correlated with the latent payoff shifters $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_J)$, but the nature of this dependence is left unspecified. The econometrician observes $Y \in \mathcal{Y}$, $X \in \mathcal{X}$, and instrumental variables $Z \in \mathcal{Z}$, which are statistically independent of ε.

Note that X is endogenous and its data-generating process is left unspecified. Such X can be viewed as part of the outcome vector (Y, X). Denote $v_{jk} = v_j(x_k)$, for all (j, k), and let $\theta = ((v_{jk})_{j=1}^J)_{k=1}^K$; denote $U_j \equiv \varepsilon_j - \varepsilon_0$, for all j, and let $U = (U_1, \dots, U_J) \in \mathbb{R}^J$. Then, given U and θ, the model produces a set of possible values for (Y, X) given by

$$G(U; \theta) = \{(y_j, x_k) : v_{jk} - v_{lk} \geqslant U_l - U_j \text{ for all } l \neq j\}.$$

Figure 7 illustrates possible realizations of $G(U; \theta)$ for some fixed θ in a model with $\mathcal{Y} = \{y_0, y_1, y_2\}$ and $X \in \{x_1, x_2\}$, assuming that $v_{11} < v_{12}$ and $v_{21} > v_{22}$. Dashed lines outline the partition of the latent variable space that corresponds to possible realizations of $G(U; \theta)$,

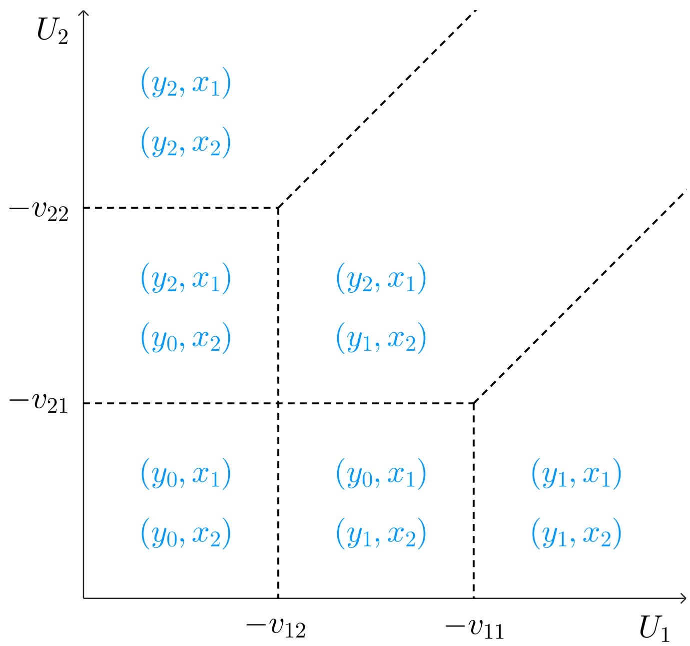


Figure 7: Set-valued predictions in a discrete choice model from Example 4 with $J = K = 2$, assuming that $v_{11} < v_{12}$ and $v_{21} > v_{22}$.

highlighted in blue.

Figure 8 depicts the corresponding bipartite graph. The upper part represents the outcome space, $\mathcal{Y} \times \mathcal{X}$, and the lower part corresponds to the partition of latent variable space in Figure 7. For example, $u_4 = \{(U_1, U_2) : U_1 \leqslant -v_{11}, U_2 \leqslant -v_{22}\}$. Depending on the values of $\theta = (\{v_{jk}\}_{j,k}, \gamma)$, the partition and the probabilities of the corresponding regions differ, but as long as $v_{11} < v_{12}$ and $v_{21} > v_{22}$, the corresponding bipartite graph remains the same. Suppose that all $\theta \in \Theta$ satisfy this restriction.[19] Then, the smallest CDC does not change with θ or Z, so it only needs to be computed once. Since $P(G(U;\theta) \subseteq A)$ does not depend on z, the sharp identified set is given by

$$\Theta_0 = \{\theta \in \Theta : \operatorname*{ess\,inf}_{z \in \mathcal{Z}} P((Y, X) \in A \,|\, Z = z) \geqslant P(G(U;\theta) \subseteq A) \text{ for all } A \in \mathcal{C}^*\}.$$

If $X \in \{x_1, \ldots, x_K\}$, the power set of the outcome space grows proportionally to $2^{(J+1)K}$. Yet, due to the simple structure of the underlying bipartite graph, the smallest CDC appears to grow proportionally to 2^K. Table 5 summarizes the results for $K \in \{2, \ldots, 15\}$.

The analysis above is similar to Chesher et al. (2013): They also treat X as part of the outcome vector and condition only on Z, which leaves $F_{U|X=x}$ completely unspecified. The inequalities in $\mathcal{C}^*$ coincide with those obtained by Chesher et al. (2013), yet our results

[19] Otherwise, partition the parameter space as in Example 1 in the main text.

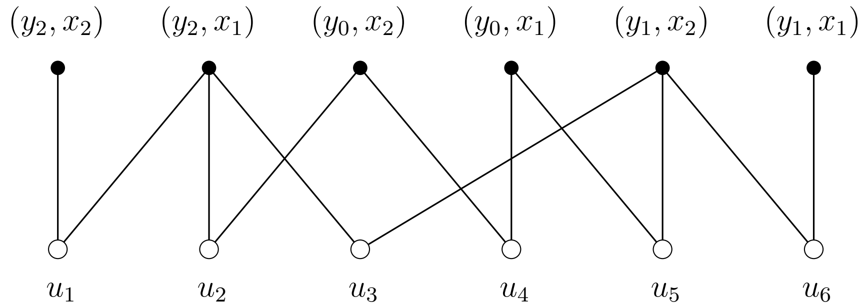


Figure 8: Discrete choice model from Example 4 with $J = 2$ and $X \in \{x_1, x_2\}$.

K	2	3	4	5	6	7	8
Total	62	510	4,094	32,766	$0.2 \cdot 10^6$	$2 \cdot 10^6$	10^7
Smallest	12	33	82	188	406	842	1,703
K	9	10	11	12	13	14	15
Total	10^8	10^9	10^{10}	10^{11}	10^{11}	10^{12}	10^{13}
Smallest	3,397	6,733	13,321	26,372	52,298	103,912	206,828

Table 5: Core-determining classes in the discrete choice model from Example 4.

additionally imply that the characterization of Θ_0 cannot be further simplified, without loosing sharpness. Tebaldi et al. (2019) take a different approach. They introduce the Minimal Relevant Partition (MRP), which is conceptually similar to the partition in Figure 7, and condition on both X and Z, treating the probabilities that the conditional distribution $F_{U|X=x}$ assigns to each of the regions in MRP, denoted $\eta = (\eta_1, \ldots, \eta_{|MRP|})$, as unknown parameters. Theorem 2.33 in Molchanov and Molinari (2018) implies that the two approaches are equivalent and deliver the same sharp identified sets. If the functional of interest depends only on η and Z is discrete, the MRP offers substantial computational advantages. If the support of X is relatively small, but the support of Z is very rich, the CDC approach may be computationally simpler. ∎

The final example revisits the network formation model of Gualdani (2021).

Example 5 (Directed Network Formation)**.** N firms form directed links with each other. The strategy of each firm is a binary vector $Y_j = (Y_{jk})_{k \neq j} \in \{0,1\}^{N-1}$, where Y_{jk} indicates the presence of a directed link from j to k, and the outcome of the game is $Y \in \{0,1\}^{N(N-1)}$. The solution concept is Pure Strategy Nash Equilibrium (PSNE). Since the total number of directed networks with N players is $2^{N(N-1)}$, the size of the outcome space $\mathcal{Y}$ of this game is $2^{2^{N(N-1)}}$. This renders sharp identification practically infeasible, even for small N. To simplify the analysis and motivate inequality selection, Gualdani (2021) imposes further restrictions on the model. The discussion below is conditional on covariates $X = x$.

First, for each firm k, define a *local game* Γ_k in which the remaining $N-1$ firms decide whether to form a directed link to firm k. Let $Y^k = (Y_1^k, \ldots, Y_N^k) \in \mathcal{Y}^k$ denote the outcome of Γ_k. Suppose the payoff of firm j is additively separable, $\pi_j(Y, \varepsilon; \theta) = \sum_{k \neq j} \pi_j^k(Y^k, \varepsilon^k; \theta)$, where each $\pi_j^k(Y^k, \varepsilon^k; \theta)$ is the same as in the entry game in Example 1 with $\delta_j > 0$. Then, the payoff from each local game depends only on the outcome of that local game, and Y is a PSNE if and only if Y^k is a PSNE of Γ_k, for all k. Second, suppose that the local games are statistically independent — that is, both $\varepsilon^1, \ldots, \varepsilon^N$ and the corresponding selection mechanisms are mutually independent.

Under the above assumptions, the random set of equilibria of the game $G(\varepsilon)$ is a Cartesian product of N independent random sets $G^k(\varepsilon^k)$ of equilibria in the local games. It follows that $\text{Core}(G^1) \times \cdots \times \text{Core}(G^N) = \text{Core}(G) \cap \mathcal{S}$, where $\mathcal{S}$ is the set of distributions on $\mathcal{Y}$ with independent marginals over $\mathcal{Y}^k$. If the distribution of the data lies in $\mathcal{S}$, the identified sets

$$\Theta_0 = \{\theta \in \Theta : P(Y \in A) \geqslant P(G \subseteq A) \; \forall A \subseteq \mathcal{Y}\};$$

$$\Theta_0' = \{\theta \in \Theta : P(Y^k \in A^k) \geqslant P(G^k \subseteq A^k) \; \forall A^k \subseteq \mathcal{Y}^k, \; \forall k\}$$

are equal. If the distribution of the data does not lie in $\mathcal{S}$, then $\Theta_0 \subseteq \Theta_0'$, because the latter checks a subset of inequalities from the former. To characterize Θ_0', Theorem 1 can be applied to each Γ_k separately. For $N = 3$, there are 254 inequalities in total and 15 in the smallest class. For $N = 4$, there are 10^{19} inequalities in total and only 144 in the smallest class. For $N = 5$, there are 10^{307} inequalities in total and 95,080 in the smallest class. Although the computational burden is lifted substantially, the resulting set of inequalities is still too large. To this end, one can adopt a type-heterogeneity assumption as in Example 1 in the main text to keep the analysis tractable. The details are left for future research. ■