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: February 19, 2025

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 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 non-redundant 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 simplify computation without losing sharpness. 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, Russell, and Stringham, 2022); network formation models (e.g., Miyauchi, 2016; De Paula, Richards-Shubik, and Tamer, 2018; Sheng, 2020; Gualdani, 2021); English auctions (e.g., Haile and Tamer, 2003; Aradillas-López, Gandhi, and Quint, 2013); models with missing or interval data (e.g., Manski and Sims, 1994; Manski, 2003; Beresteanu, Molchanov, and Molinari, 2011); potential outcome models (e.g., Heckman, Smith, and Clements, 1997; Manski and Pepper, 2000, 2009; Beresteanu, Molchanov, and Molinari, 2012; Russell, 2021); and discrete choice models with endogeneity (e.g., Chesher, Rosen, and Smolinski, 2013; Chesher and Rosen, 2017; Torgovitsky, 2019; Tebaldi, Torgovitsky, and Yang, 2019) or unobserved or counterfactual choice sets (e.g., Manski, 2007; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021).

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

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

must hold for all $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$. So, 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) imply that the inequalities in (1) hold if and only if $Y \in G(U, X; \theta)$, almost surely. Thus, assuming the parameter space Θ captures all other restrictions imposed on the model, the identified set Θ_0 is sharp.

The above characterization 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 of the examples cited below, the set-valued predictions naturally arise in the space of latent variables: 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 $\theta_0 \in \Theta_0$. The analysis in this paper applies symmetrically in such 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, Li, and Mourifié, 2020).

At the same time, examples suggest that many of the inequalities in (2) may be redundant, in the sense that excluding them from the analysis 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 an 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 its' computational complexity is proportional to the size of the smallest CDC. When the smallest CDC is still infeasible, we discuss imposing additional assumptions to motivate further inequality selection without losing sharpness. 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, Porter, Ho, and Ishii (2015); Molinari (2020); Chesher and Rosen (2020); and Kline, Pakes, and Tamer (2021) for detailed reviews. The key object in the identification analysis is the set $\mathcal{P}(x; \theta)$ of 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 $\mathcal{P}(x; \theta)$.

The most closely related papers are Galichon and Henry (2011); Chesher and Rosen (2017); and Luo and Wang (2018). As in this paper, those authors represent the set $\mathcal{P}(x; \theta)$ using the 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 itin practice. In turn, Chesher and Rosen (2017) derive analytical sufficient conditions for identifying redundant Artstein's inequalities. In this paper, we obtain a set of necessary and sufficient conditions for redundancy and use it to characterize the smallest possible CDC.

Luo and Wang (2018) also provide 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 fact 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, conditional on an excluded instrumental variable, can be intersected, 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, Section 5 extends the main results to settings in which the outcome variable has infinite support.

Other closely related papers are Beresteanu, Molchanov, and Molinari (2011) and Mbakop (2023). Beresteanu, Molchanov, and Molinari (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 (see also Pakes and Porter, 2024). We argue that the CDC approach is complementary to these methods and enables faster computation of the sharp identified set and simpler inference procedures in many settings.

Other related work includes Tebaldi, Torgovitsky, and Yang (2019) and Gu, Russell, and Stringham (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 cells 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 the cell 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 rest of the paper is organized as follows. Section 2 presents motivating examples and provides the necessary background. Section 3 presents novel theoretical results. Section 4 provides an algorithm to compute the smallest core-determining class and compares the proposed approach with other methods. Section 5 provides an extension to models in which the outcomes have infinite support. Section 6 illustrates the utility of selecting inequalities, and 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 featuring discrete-outcome models. Additional examples are considered in Section 3.3 and Appendix C, 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, Molchanov, and Molinari (2011); and Aradillas-López (2020).

Example 1 (Static Entry Game)**.** Each of N firms, indexed by $j = 1, \ldots, N$, 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

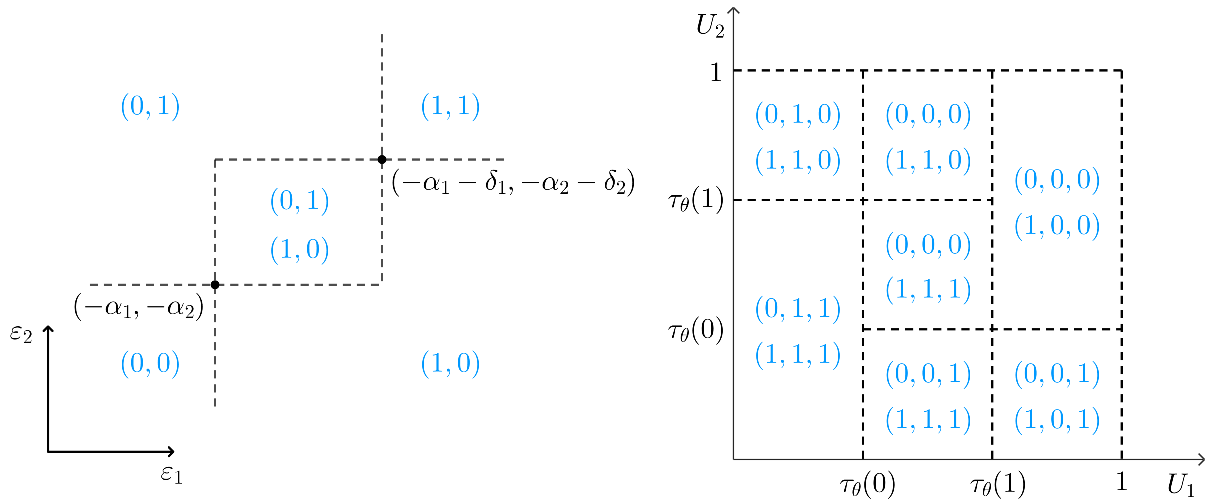
$$\pi(X_t, A_t, \varepsilon_t) = \begin{cases} \bar{\pi} - \varepsilon_t & \text{if } X_t = 1, A_t = 1; \\ \bar{\pi} - \varepsilon_t - \gamma & \text{if } X_t = 0, A_t = 1; \\ 0 & \text{otherwise,} \end{cases}$$

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 $Y = (X_1, A_1, \dots, A_T) \in \{0,1\}^{T+1}$.

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



(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.

conditions, there is a unique stationary solution

$$A_t = \mathbf{1}(U_t \leqslant \tau_\theta(X_t)),$$

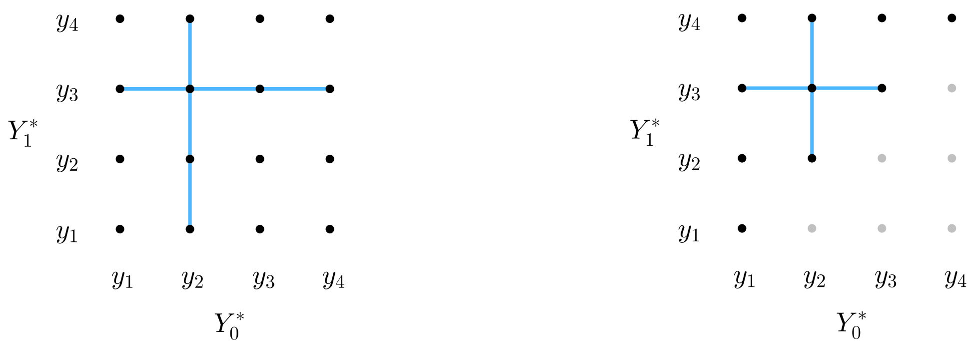
where U_t is the quantile transformation of ε_t, and τ is an increasing function of X_t known up to the parameters $\theta = (\bar{\pi}, \gamma, \rho)$.

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 $Y = (X_1, A_1, \dots, A_T)$. Then, given $U = (U_1, \dots, U_T)$ and θ, the model produces a set of possible values for Y given by

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

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 that has been studied by Balke and Pearl (1997); Heckman, Smith, and Clements (1997); Heckman and Vytlacil (2007); Beresteanu, Molchanov, and Molinari (2012); Lee and Salanié (2018); Heckman and Pinto (2018); Mourifie, Henry, and Meango (2020); Russell (2021); and Bai, Huang, Moon, Shaikh, and Vytlacil (2024), among many others.



(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 $|\mathcal{Y}| = 4$.

Example 3 (Potential Outcomes Models)**.** Let $D \in \mathcal{D}$ denote the treatment assignment, $Y^* = (Y_d^*)_{d\in\mathcal{D}} \in \mathcal{Y}^{|\mathcal{D}|}$ — potential outcomes, $Y = \sum_{d\in\mathcal{D}} Y_d^* \mathbf{1}(D = d) \in \mathcal{Y}$ — 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 $Y^* \in \mathcal{S}_{Y^*}$ for some known set $\mathcal{S}_{Y^*} \subseteq \mathcal{Y}^{|\mathcal{D}|}$ (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 straightforward 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 $Y_{d'}^*$ for $d' \neq d$ is that $Y_{d'} \in \mathcal{Y}$ and $Y^* \in \mathcal{S}_{Y^*}$. Thus, the set-valued prediction for Y^* can be written as:

$$G(Y, D) = \sum_{d\in\mathcal{D}} \mathbf{1}(D = d) B_d(Y) \cap \mathcal{S}_{Y^*},$$

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 below and refer the reader to Molchanov and Molinari (2018) for a textbook treatment.

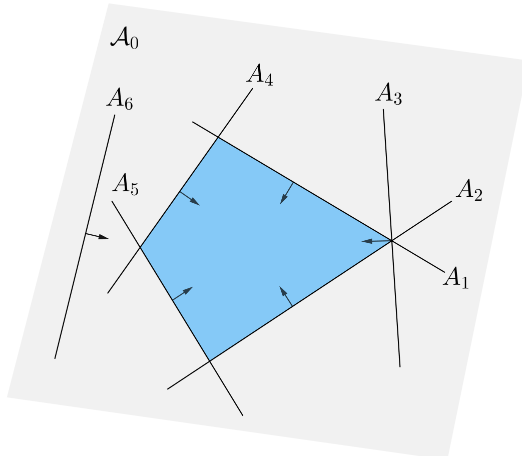


Figure 3: The core of a random set.

Until Section 5, we focus on settings in which the outcome space $\mathcal{Y}$ is finite.

Let $(\mathcal{U}, \mathcal{F}, P)$ be a complete probability space and $(\mathcal{Y}, \mathcal{B})$ — a finite measurable space, with $\mathcal{Y} = \{y_1, \ldots, y_S\} \subseteq \mathbb{R}^{d_\mathcal{Y}}$ and $\mathcal{B} = 2^\mathcal{Y}$. Let $\mathcal{M}$ denote the set of all probability measures on $(\mathcal{Y}, \mathcal{B})$. Let $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ be a correspondence. For each $A \in \mathcal{B}$, denote the upper and lower inverse of G by

$$
\begin{aligned}
G^-(A) &= \{u \in \mathcal{U} : G(u) \subseteq A\}; \\
G^{-1}(A) &= \{u \in \mathcal{U} : G(u) \cap A \neq \varnothing\},
\end{aligned}
\tag{3}
$$

and note that $G^-(A) \subseteq G^{-1}(A)$. If the correspondence G is measurable, in the sense that $G^-(A) \in \mathcal{F}$ for all $A \in \mathcal{B}$, it defines a *random (closed) set*. Its distribution can be described by the *containment functional*, defined for all $A \in \mathcal{B}$ as

$$C_G(A) = P(G \subseteq A).$$

The *support* of a random set G, denoted $\mathcal{S} \subseteq \mathcal{B}$, is the collection of sets $A \in \mathcal{B}$ such that $P(G = A) > 0$. Any random variable $Y : (\mathcal{U}, \mathcal{F}, P) \to (\mathcal{Y}, \mathcal{B})$ that satisfies $P(Y \in G) = 1$ is called a *selection* of G. The set of distributions of all selections is called the *core*, and will be denoted $\mathrm{Core}(G)$. Artstein (1983) showed that the core consists of all probability distributions that dominate the containment functional:

$$\mathrm{Core}(G) = \{\mu \in \mathcal{M} : \mu(A) \geqslant C_G(A) \text{ for all } A \in \mathcal{B}\}. \tag{4}$$

The inequalities in (4) are known as Artstein's inequalities. To fully characterize the core, it usually suffices to consider smaller classes of sets.

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

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

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

Two types of sets will play an important role in the analysis below.

Definition 2.2 (Critical and Implicit Equality Sets)**.** *A set $A \in \mathcal{B}$ is critical if $\mathcal{M}(\mathcal{B}\backslash\{A\}) \neq \mathcal{M}(\mathcal{B})$. A set $A \in \mathcal{B}$ is an implicit equality set if $\mu(A) = C_G(A)$ for all $\mu \in Core(G)$.*

Any core-determining class must contain all critical sets and ensure that all implicit equality constraints hold. Figure 3 provides a stylized illustration in $\mathcal{M}$. Here, $\mathcal{A}_0$ denotes the class of all implicit equality sets, and the gray shaded region depicts the set $\{\mu \in \mathcal{M} : \mu(A) = C_G(A) \text{ for all } A \in \mathcal{A}_0\}$. Each straight line corresponds to an Artstein's inequality with an arrow that indicates the direction in which it is 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.

2.3 Identifying Redundant Inequalities

To construct a core-determining class, it is necessary to understand the implications between Artstein's inequalities. Specifically, for what triplets of sets $A_1, A_2, A \in \mathcal{B}$, do the inequalities $\mu(A_1) \geqslant C_G(A_1)$ and $\mu(A_2) \geqslant C_G(A_2)$ imply $\mu(A) \geqslant C_G(A)$, for all $\mu \in \text{Core}(G)$? In situations in which the containment functional is additive, the answer is fairly straightforward.

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 $G^-(A_1 \cup A_2) = G^-(A_1) \cup G^-(A_2)$. The third 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, given $\mu(A_1) \geqslant C_G(A_1)$ and $\mu(A_2) \geqslant C_G(A_2)$,

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

so A is redundant given A_1 and A_2.[2]

[2] As a special case, consider a set A that cannot be expressed as a union of elements of the support of G, i.e., $A \neq G(G^-(A))$, where $G(G^-(A)) = \bigcup_{\omega\in\Omega}\{G(\omega) : G(\omega) \subseteq A\}$. Then, setting $A_1 = G(G^-(A))$ and $A_2 = A\backslash A_1$, it follows that $G^-(A_1) = G^-(A)$ and $G^-(A_2) = \varnothing$. Therefore, given $\mu(A_1) \geqslant C_G(A_1)$, we have $\mu(A) \geqslant \mu(A_1) \geqslant C_G(A_1) = C_G(A)$, so A is redundant given A_1. Thus, one may restrict attention to sets A which can be expressed as unions of elements of the support. See the errata to Beresteanu, Molchanov, and Molinari (2012), Chesher and Rosen (2017), and Theorems 2.22–2.23 in Molchanov and Molinari (2018) for related arguments.

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 third 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, given $\mu(A_1) \geqslant C_G(A_1)$ and $\mu(A_2) \geqslant C_G(A_2)$,

$$1 + \mu(A) = \mu(A_1) + \mu(A_2) \geqslant C_G(A_1) + C_G(A_2) = 1 + C_G(A), \tag{6}$$

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 $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \varnothing$, which will be useful in the sequel.

Below, we show that no other non-trivial implications between Artstein's inequalities exist. We use this fact to characterize all critical and implicit equality sets analytically and provide an efficient algorithm to obtain the smallest possible core-determining class.

3 The Smallest Core-Determining Class

Suppose the model postulates 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[3]

$$\Theta_0 = \{\theta \in \Theta : P_{Y|X=x}(A) \geqslant C_{G(U,x;\theta)}(A), \text{ for all } A \in \mathcal{C}(x, \theta), \text{ a.s. } x \in \mathcal{X}\}, \tag{7}$$

where $\mathcal{C}(x, \theta) \subseteq \mathcal{B}$ is a core-determining class for the random set $G(U, x; \theta)$ conditional on $X = x$. In this section, we characterize the smallest possible core-determining class $\mathcal{C}^*(x; \theta)$, clarify how it depends on x and θ, and characterize sharp identified sets in a tractable way in several popular applications.

3.1 Random Sets as Bipartite Graphs

The first step is to represent the random set $G(U, x; \theta)$, conditional on $X = x$, as a bipartite graph. Let $S(x, \theta) = \{G_1, \ldots, G_K\}$ denote the support of $G(U, x; \theta)$ conditional on $X = x$, i.e., the set of all values $G_k \subseteq \mathcal{Y}$ such that $P(G(U, x; \theta) = G_k \,|\, X = x) > 0$. Partition the latent variable space $\mathcal{U}$ as $\mathcal{U}(x, \theta) = \{u_1, \ldots u_K\}$, where $u_k = \{u \in \mathcal{U} : G(u, x; \theta) = G_k\}$, and 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 equivalently defined as a correspondence $G : (\mathcal{U}(x,\theta), 2^{\mathcal{U}(x,\theta)}, P_{(x,\theta)}) \rightrightarrows \mathcal{Y}$ between two finite spaces. Such correspondence can represented by an undirected bipartite graph $\mathbf{B}$ with vertices $V(\mathbf{B}) = (\mathcal{U}, \mathcal{Y})$ and edges

[3] The representations via unconditional and conditional Artstein's inequalities are equivalent; see Theorem 2.33 in Molchanov and Molinari (2018).

$E(\mathbf{B}) = \{(u, y) \in \mathcal{U} \times \mathcal{Y} : y \in G(u)\}$.

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 $G^{-1}(\{(0, 1, 1), (1, 1, 1), (0, 0, 0)\}) = \{u_1, u_2, u_3, u_5, u_6\}$.

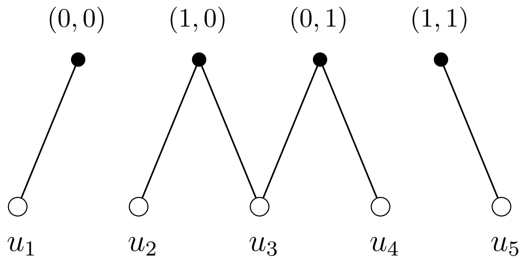
Panel (c) depicts the potential outcomes model from Example 3 with $\mathcal{D} = \{0, 1\}$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and $\mathcal{S}_{Y^*} = \mathcal{Y}^2$. The upper part is $\mathcal{S}_{Y^*}$, and the lower part is $\mathcal{D} \times \mathcal{Y}$. For example, $G((0, 2)) = \{(2,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

For any given x and θ, the bipartite graph $\mathbf{B}$ can easily 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 until there is a discrete "regime change." Sections 3.3 below provides detailed examples.

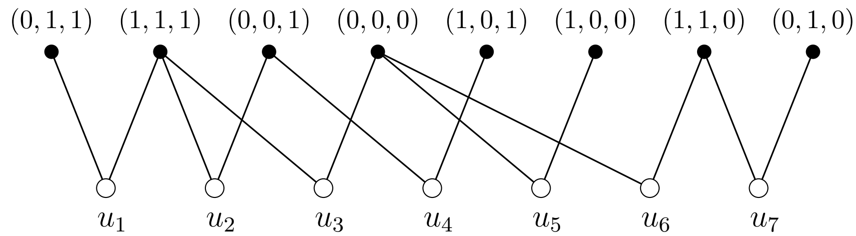
3.2 The Smallest Core-Determining Class

The implications between Artstein's inequalities discussed in Section 2.3 can be expressed in terms of the connectivity of suitable subgraphs of $\mathbf{B}$. A subgraph of $\mathbf{B}$ induced by the vertices $(V_\mathcal{Y}, V_\mathcal{U})$ is an undirected bipartite 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.

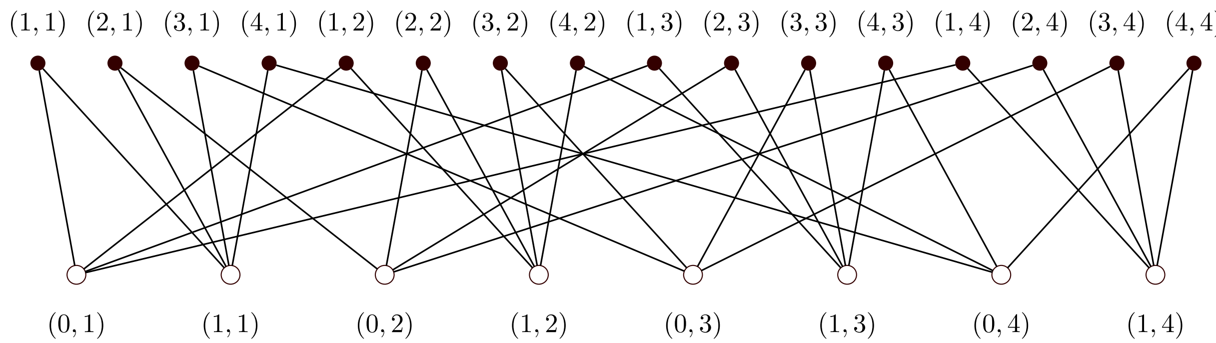
For example, Consider the graph in Panel (b) of Figure 4. First, let $A_1 = \{(0, 1, 1), (1, 1, 1)\}$, $A_2 = \{(1, 1, 0), (0, 1, 0)\}$, and $A = A_1 \cup A_2$. Then, $G^-(A_1) = \{u_1\}, G^-(A_2) = \{u_7\}$, and $G^-(A) = \{u_1, u_7\}$. Thus, A is redundant given A_1 and A_2, as in Equation (5). Note that in this case, the subgraph induced by $(A, G^-(A))$ is disconnected. Second, let



(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.

$A = \{(0, 0, 1), (0, 0, 0), (1, 0, 1)\}$ and $A_1 = \{(0, 0, 1), (1, 0, 1)\} \subset A$. Such 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 in Equation (**??**). In this case, again, the subgraph induced by $(A, G^-(A))$ is disconnected. Finally, let $A = \{(1, 1, 1), (0, 0, 1), (0, 0, 0)\}$, $A_1 = A \cup \{(0, 1, 1)\}$, $A_2 = A \cup \{(1, 0, 1), (1, 0, 0), (1, 1, 0), (0, 1, 0)\}$, 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 (6). In this case, the subgraph induced by $(A^c, G^{-1}(A^c))$ is disconnected.

Thus, for any redundant set $A \subseteq \mathcal{Y}$ identified by (5)–(6), the subgraph of $\mathbf{B}$ induced by either $(A, G^-(A))$ or by $(A^c, G^{-1}(A^c))$ is disconnected. Conversely, it turns out that if both subgraphs are connected, the set A must be critical.

Lemma 1. *Let $\mathcal{U} = \{u_1, \dots, u_K\}$, $\mathcal{Y} = \{y_1, \dots, y_S\}$, 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 $k = 1, \dots, K$. Then, a set $A \subseteq \mathcal{Y}$ is critical if and only if the subgraphs of $\mathbf{B}$ induced by $(A, G^{-}(A))$ and $(A^c, G^{-1}(A^c))$ are connected.*

The proof of this result is constructive: Given a set A that satisfies the above assumptions, 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 A corresponds to one of the faces of the convex polytope representing the Core(G), as in Figure 3, which in turn implies that A is critical. The assumption $P(u_k) > 0$ for all $k = 1, \dots, K$ merely ensures that there are no redundant elements in $\mathcal{U}$. If $P(u_k) = 0$, then u_k can simply be removed from $\mathcal{U}$ and $\mathbf{B}$, together with all its edges. In turn, the assumption that $\mathbf{B}$ is connected is substantive and related to implicit equality sets.

Lemma 2. *Let $\mathcal{U} = \{u_1, \dots, u_K\}$, $\mathcal{Y} = \{y_1, \dots, y_S\}$, and $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ be a non-empty random set with a bipartite graph $\mathbf{B}$. Let $\mathcal{Y} = \bigcup_{l=1}^{L} \mathcal{Y}_l$ denote 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, $A \subseteq \mathcal{Y}$ is an implicit equality set if and only $A = \bigcup_{l \in L_A} \mathcal{Y}_l$ for some $L_A \subseteq \{1, \dots, L\}$.*

In particular, Lemma 2 shows that whenever implicit equality sets exist, there are at least two of them and that none of the implicit equality sets is critical. In the setting of Lemma 2, the bipartite graph $\mathbf{B}$ "breaks" into L connected components $\mathbf{B}_l$ with vertices $V(\mathbf{B}_l) = (G^{-1}(\mathcal{Y}_l), \mathcal{Y}_l)$ 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 (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 the insights of Lemmas 1 and 2 yields a simple characterization of the smallest possible CDC.

Theorem 1. *Let $\mathcal{U} = \{u_1, \dots, u_K\}$, $\mathcal{Y} = \{y_1, \dots, y_S\}$, and $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ be a non-empty random set with a bipartite graph $\mathbf{B}$. Suppose $P(u_k) > 0$ for all $k = 1, \dots, K$. Then*

1. *If $\mathbf{B}$ is connected, the class of all critical sets characterized in Lemma 1 is the smallest core-determining class.*

2. *If $\mathbf{B}$ can be decomposed into connected components, as in Lemma 2, there are L core-determining classes of the same, smallest possible cardinality. Specifically, letting $\mathcal{C}_l^*$ denote the class of all critical sets in $\mathbf{B}_l$, characterized in Lemma 1, the class $\mathcal{C}^* = \bigcup_{j=1}^{L} \mathcal{C}_j^* \cup \bigcup_{j \neq l} \mathcal{Y}_j$ is core-determining, for each $l = 1, \dots, L$.*

This result has two key implications. The first one is stated as a corollary.

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

As Gu, Russell, and Stringham (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 $S(x;\theta) = S_m(x)$ for all $\theta \in \Theta_m$, for each $m \in \{1, \dots, M\}$. 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 $S(x;\theta) = S(x';\theta)$ for all $x, x' \in \mathcal{X}$, for 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 allows us to devise an algorithm that avoids checking all $2^{|\mathcal{Y}|} - 2$ candidate inequalities for redundancy. First, the algorithm decomposes $\mathbf{B}$ into connected components to obtain implicit equality sets. Second, the algorithm "builds up" all critical sets iteratively within each connected component. The worst-case complexity of the algorithm is $|\mathcal{C}^*|(|\mathcal{Y}| + |\mathcal{U}| + |\mathcal{E}|)$, where $|\mathcal{C}^*|$ is the size of the smallest CDC, $|\mathcal{Y}| + |\mathcal{U}|$ is the number of vertices, and $|\mathcal{E}|$ is the number of edges in $\mathbf{B}$. Details are provided in Section 4.

3.3 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 or equilibrium selection 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.[4] 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. $x \in \mathcal{X}$. Then, assuming also that $U = (\varepsilon_1, \ldots, \varepsilon_N)$ and X are statistically independent, the sharp identified set for θ can be expressed as

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

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.[5] 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.

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. This fact renders the corresponding bipartite graph very interconnected, which complicates identification. 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 m. 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.

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

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

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.[6] 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 new 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. Extension to three or more types is straightforward. ■

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

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

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, which are 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. ■

[6] 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, Molchanov, and Molinari (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	31	49	344

(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.

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, Z is is independent of Y^*,

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

where $\mathcal{C}$ denotes the smallest possible core-determining class.[7]

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

[7] 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 simpler. See Section 4.3 for a related discussion.

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.

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 $\mathcal{S}_{Y^*}^I = \{y^* \in \mathcal{Y}^{|\mathcal{D}|} : y_d^* \leqslant y_{d+1}^* \text{ for all } d = 1, \ldots, |\mathcal{D}| - 1\}$ ensures that $d \mapsto Y_d^*$ is increasing and $\mathcal{S}_{Y^*}^{IC} = \mathcal{S}_{Y^*}^I \cap \{y^* \in \mathcal{Y}^{|\mathcal{D}|} : y_{d+1}^* - y_d^* \geqslant y_{d+2}^* - y_{d+1}^* \text{ for all } d = 1, \ldots, |\mathcal{D}| - 2\}$ further imposes that $d \mapsto Y_d^*$ is concave. These assumptions substantially restrict the outcome space and the corresponding bipartite graphs, which results in a much smaller CDC. Panels (b) and (c) of Table 2 illustrate.

Finally, consider imposing more structure on the relationship between D and Z. Suppose that in addition to the vector of potential outcomes Y^*, 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^*)

$$G(Y, D, Z) = \left\{\sum_{d \in \mathcal{D}} \mathbf{1}(D = d) B_d(Y) \times \sum_{z \in \mathcal{Z}} \mathbf{1}(Z = z) B_z(D)\right\} \cap \mathcal{S},$$

where $B_d(Y) = (\mathcal{Y} \times \cdots \times \{Y\} \times \ldots \mathcal{Y})$ with $\{Y\}$ in the d-th component, and $B_z(D) = (\mathcal{D} \times \cdots \times \{D\} \times \ldots \mathcal{D})$ 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 $G(y, d, z)$. The Artstein's inequalities reduce to equalities of the form $P(Y_d^* = y, D_z^* = d) = P(Y = y, D = d \,|\, Z = z)$, for all $(y, d) \in \mathcal{S}$, $z \in \mathcal{Z}$. These equalities also follow directly from the assumed relationships between (Y, D, Z) and (Y^*, D^*), and independence of the instrument, as in Balke and Pearl (1997) or Bai, Huang, Moon, Shaikh, and Vytlacil (2024). ■

4 Implementation and Relation to Other Methods

4.1 The Master Algorithm

Algorithm 1 below summarizes all the steps necessary to characterize the sharp identified set Θ_0 as in Equation (7). Throughout, we assume that X is discrete or have 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, Russell, and Stringham (2022). (**Note:** this step is not always required, as discussed in detail in Section 3.3.)

2. **Partition the latent variable space**. Fix $m \in \{1, \ldots, M\}$ and any $\theta \in \Theta_m$. Let $\mathcal{Y} = \{y_1, \ldots, y_S\}$ denote the outcome space and $S(x; \theta) = \{G_1, \ldots, G_K\}$ denote the support of $G(U, x; \theta)$, conditional on $X = x$. Partition the latent variable space as $\mathcal{U}(x, \theta) = \{u_1, \ldots u_K\}$, 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 $k = 1, \ldots, K$. The probabilities $P_{(x,\theta)}$ can be computed by resampling or numerical integration.

3. **Construct the bipartite graph.** Define vertices $v_1, \ldots, v_S$ corresponding to $\mathcal{Y}$ and $v_{S+1}, \ldots, v_{S+K}$ corresponding to $\mathcal{U}(x;\theta)$. Define the edges (v_{S+k}, v_l) for all $v_l \in G_k$, for all $k = 1, \ldots, K$. 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 $m = 1, \ldots, M$ 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 simply stacked together. 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 depend on x, so partitioning Θ as in Step 1 is not required. Then, depending on the setting, the above conditional inequalities may either be intersected over X or turned into unconditional inequalities using some instrument functions $h : \mathcal{X} \to \mathbb{R}^d_+$ via $\mathbb{E}[(\mathbf{1}(Y \in A) - \mathbf{1}(G(U; X; \theta) \subseteq A)h(X)] \geqslant 0$. Importantly, such transformation looses sharpness and may lead to discordant relaxations as discussed in Kédagni, Li, and Mourifié (2020), unless $d \to \infty$.

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 easily be found by decomposing the graph $\mathbf{B}$ into connected components, so the main challenge is to locate the critical sets within each connected component. To simplify notation, suppose that the graph $\mathbf{B}$ itself is connected. Say 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. Also, say that a set C is a *minimal critical superset* of A if there is no critical set $\tilde{C}$ such that $A \subset \tilde{C} \subset C$.

The idea is to 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 $F(\mathcal{C}) = \cup_{A \in \mathcal{C}} F(A)$. Our proposed algorithm iterates 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 discover 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 $Q = \{A \cup G(u) : u \in G^{-1}(A) \backslash G^{-}(A)\}$.

2. For each $C \in Q$:

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

   - Collect all sets of the form $C \cup \bigcup_{j \neq l} \mathcal{Y}_j$ 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 $\mathcal{Y}_l$ parts of these components because all other configurations cannot be complement-connected.

Then, the smallest CDC can be 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 $k = 1, \ldots, K$:

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

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

   - 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 R_k.

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

Formal proofs are provided in Appendix A.4. Since at every iteration — except possibly the first — Algorithm 2 is only applied to critical sets, the worst-case complexity of Algorithm 3 is proportional to the number of critical sets times the cost of decomposing subgraphs of

$\mathbf{B}$ into connected components. The time complexity of decomposing the whole graph $\mathbf{B}$ into connected components using Depth First Search is $|\mathcal{Y}| + |\mathcal{U}| + |\mathcal{E}|$, where $|\mathcal{Y}| + |\mathcal{U}|$ and $|\mathcal{E}|$ are the numbers of vertices and edges in $\mathbf{B}$ correspondingly.[8] Therefore, the worst-case time complexity of Algorithm 3 is of order $\max(|CDC|, |\mathcal{U}|) \times (|\mathcal{Y}| + |\mathcal{U}| + |\mathcal{E}|)$, where $|CDC|$ is the size of the smallest CDC.

Algorithm 3 can be efficiently implemented in any programming language that has a native implementation of sets (e.g., Python or Julia). Since the algorithm essentially only looks at the critical sets, it is able to compute the smallest CDC quickly whenever it is tractable. For example, with the Julia implementation, in all examples considered in Section 3.3 in which the CDC has cardinality less than 1,000, computation takes at most several seconds, even in settings in which the total number of inequalities is prohibitively large and other algorithms are infeasible. Since the complexity is proportional to the size of the smallest CDC, further substantial improvements are not possible.

4.3 Comparison with Other Approaches: Additional Restrictions, Counterfactuals, and Inference

Besides Artstein's inequalities, several alternative approaches are available for characterizing sharp identified sets in models with set-valued predictions. This section describes each approach in more detail and compares it with the CDC approach in terms of (i) computational tractability; (ii) obtaining sharp bounds counterfactual quantities $\phi(\theta_0)$; and (iii) inference.

Recall that $\mathcal{P}(x;\theta)$ denotes the set of distributions of the outcome Y, given covariates $X = x$ and a parameter value $\theta \in \Theta$, predicted by the model. 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 all of its faces. When the smallest CDC is tractable, the Artstein's inequalities approach provides a tractable characterization of the sharp identified set Θ_0 and has several attractive features.

First, as illustrated in Section 3.3, additional restrictions on the model — such as instrument exogeneity, support restrictions, and restrictions on the underlying selection mecha-

[8] See, e.g., Section 3.2 in Kleinberg and Tardos (2006) for the detailed discussion.

nisms — 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. Assuming that Θ_0 is a connected set and $\theta \mapsto \phi(\theta)$ 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 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, Russell, and Stringham (2022), is discussed in the next section.

Third, given a collection of Arstein's moment inequalities, inference on θ_0 or its subvectors is a well-studied problem. When the smallest CDC does not change with θ, standard inference procedures for moment inequalities apply; see Canay and Shaikh (2017) for a review. A minor complication arises when the CDC changes with θ. In such settings, the parameter space is partitioned into a finite number of disjoint parts $\Theta = \bigcup_{m=1}^{M} \Theta_m$, according to the support of G, and the identified set takes the form $\Theta_0(P) = \bigcup_{m=1}^{M} \Theta_{0,m}(P)$. Letting $\hat{\phi}_{m,n}(\theta)$ denote a test for $H_{0,m} : \theta \in \Theta_{0,m}(P)$ that is valid uniformly over a set of distributions $\mathbf{P}_m$, it is easy to verify that the test $\hat{\phi}_n(\theta) = \sum_{m=1}^{M} \hat{\phi}_{m,n}(\theta)\mathbf{1}(\theta \in \Theta_m)$ for $H_0 : \theta \in \Theta_0(P)$ is valid uniformly over $\cap_{m=1}^{M} \mathbf{P}_m$, and the confidence set may be obtained by test inversion.[9] If the implicit equality sets differ across Θ_m, the above test will be more powerful than the test using all of Artstein's inequalities because it incorporates the information that certain Artstein's inequalities are binding. Existing procedures for subvector inference (see, e.g., Romano and Shaikh, 2008; Bugni, Canay, and Shi, 2017; Kaido, Molinari, and Stoye, 2019) can also be modified to accommodate situations in which the set of relevant moment inequalities depends on θ. Pursuing such modifications formally is beyond the scope of this paper.

The CDC approach identifies the inequalities that are redundant for identification. Such inequalities are also redundant for estimation of the identified set Θ_0 or bounds on any functional $\phi(\theta_0)$, provided that the plug-in estimator is appropriate.[10] 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 and

[9]The claim follows immediately from the fact that M is finite, and each of the tests $\hat{\phi}_{m,n}$ is uniformly valid, in the sense that $\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_m} \sup_{\theta\in\Theta_{0,m}(P)} \mathbb{E}_P[\hat{\phi}_{n,m}(\theta)] \leqslant \alpha$.

[10]That is, $\hat{\Theta}_n = \{\theta \in \Theta : \hat{P}_n(Y \in A \mid X = x) \geqslant P(G(U, X; \theta) \subseteq A \mid X = x) \forall A \in \mathcal{C}_\theta(x), \forall x \in \mathcal{X}\}$ without any slack in the inequalities. See Theorem 5.22 in Molchanov and Molinari (2018) for a related discussion.

existing results on admissibility of moment inequality tests suggest that the answer depends on where the researcher wants to direct the power.[11] 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, Molchanov, and Molinari (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 $s \in S$.[12] 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

$$\Theta_0 = \{\theta \in \Theta : \sup_{t\in B}(t^T\mathbb{E}[Y^*\,|\,X = x] - \mathbb{E}[h_{Q(U,x;\theta)}(t)\,|\,X = x]) \leqslant 0,\ x \in \mathcal{X} \text{ a.s.}\}. \tag{8}$$

Beresteanu, Molchanov, and Molinari (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, Molchanov, and Molinari (2012) for a related discussion. Second, Equation (8) essentially 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).

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

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

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 $P_{Y\,|\,X=x}$ 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

$$\begin{aligned} \textstyle\sum_{u\in G^{-1}(y)}\pi(y,u;x,\theta) &= P_{Y|X=x}(y) \quad &&\text{for all } y\in\mathcal{Y},\\ \textstyle\sum_{y\in G(u)}\pi(y,u;x,\theta) &= P_{(x;\theta)}(u) \quad &&\text{for all } u\in\mathcal{U}. \end{aligned} \tag{9}$$

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

$$\min_{\eta\in\mathbb{R}^{|\mathcal{Y}|+|\mathcal{U}|}}\big(b(x;\theta)^T\eta\,|\,A(x;\theta)^T\eta\geqslant 0\big)\geqslant 0, \tag{10}$$

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 (9) 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

$$\Theta_0=\{\theta\in\Theta:(10)\text{ holds } x\in\mathcal{X}\text{-a.s.}\}. \tag{11}$$

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 (9) 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 $(|\mathcal{Y}|+|\mathcal{U}|)\times|\mathcal{E}|$; see, e.g., Orlin (2013).[13]

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 instrument $Z\in\mathcal{Z}$, as in Example 3 discussed in Section 3.3.[14] 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 $|\mathcal{Z}|-1$ match-

[13] As another alternative, Galichon and Henry (2011) propose using submodular minimization. Using Artstein's inequalities, 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 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 $|\mathcal{U}|\gg|\mathcal{Y}|^3$.

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

ing 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 (10)–(11) 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, Torgovitsky, and Yang (2019) and Gu, Russell, and Stringham (2022). In Gu, Russell, and Stringham (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 $\{u_1^*, \dots, u_L^*\}$ 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, Torgovitsky, and Yang (2019), such 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 $\pi_x(y, y^*, u_l^*)$ is a valid probability distribution and $q = \mathcal{X}|\mathcal{Y}|^2 L$ non-negativity constraints.

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 $\mathcal{U}(x; \theta) = \{u_1, \dots, u_k\}$ 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 Extension: Outcomes with Infinite Support

This section extends the main theoretical results in Section 3 to models in which the outcome variable has infinite support. The main distinction here is that we have to carefully treat measure-zero sets in both latent space and outcome spaces. In view of this, the results below are somewhat more precise than those in Section 3. We start with a more general setup.

5.1 Setup

Let $(\mathcal{U}, \mathcal{F}, P)$ be a complete probability space and $(\mathcal{Y}, \mathcal{B})$ a measurable space, where $\mathcal{Y} \subseteq \mathbb{R}^d$, and $\mathcal{B}$ is the Borel σ-field of subsets of $\mathcal{Y}$. Let $\mathcal{M}$ denote the set of all probability measures on $\mathcal{B}$, and $\mathfrak{F}$ the class of all closed subsets of $\mathcal{Y}$. A *random closed set* is a measurable correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ such that $G(u) \in \mathfrak{F}$ for all $u \in \mathcal{U}$. Here, measurability requires $G^-(A) \in \mathcal{F}$ for every $A \in \mathfrak{F}$, where $G^-(A)$ is defined in (3).

As before, let Core(G) denote the set of distributions of all measurable selections of G. Given a σ-finite measure Q on $\mathcal{B}$, let $\mathcal{M}_Q = \{\mu \in \mathcal{M} : \mu \ll Q\}$ denote the set of probability measures absolutely continuous with respect to Q. For any two sets $A, B \in \mathcal{B}$ say that $A = B$, Q-a.s., if $Q((A \cap B^c) \cup (B \cap A^c)) = 0$. For any two sets $A, B \in \mathcal{F}$, define "$A = B$, P-a.s" similarly. Denote $N(A) = G^{-1}(A) \backslash G^-(A)$.

Throughout this section, we impose the following regularity condition.

Assumption 5.1 (Dominating Measure). *Let $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B})$ be a random closed set. There exists a σ-finite measure Q on $\mathcal{B}$ such that*

1. *Only the distributions $\mu \in Core(G) \cap \mathcal{M}_Q$ are of interest.*
2. *$Q(G(u)) > 0$ for P-almost all $u \in \mathcal{U}$; for all $A \in \mathcal{B}$ with $Q(A) > 0$, $Q(A \cap G(u)) > 0$ for P-almost-all $u \in N(A)$.*

Part 1 of the assumption allows us to control measure-zero sets in the outcome space $\mathcal{Y}$. In many applications, it is possible to find Q such that $\text{Core}(G) \subseteq \mathcal{M}_Q$. If the outcome $\mathcal{Y}$ is finite, Q can be taken as a counting measure on $\mathcal{Y}$; if the sets $G(u)$ have non-empty interior but can be arbitrarily narrow with positive probability, Q can be taken as a Lebesgue measure. The researcher can also choose Q to explicitly restrict the set of selections of interest. Part 2 of the assumption is a mild regularity condition that requires that the realizations G can be "detected" by the measure Q. It enables us to introduce the notion of connectivity of the correspondence G similar to that of the bipartite graph in Section 3.1.

Given Q and P, each set $A \in \mathcal{B}$ can be associated with an equivalence class $[A]$ with the relation $A' \sim A$ if $A = A'$, Q-a.s., and $G^{-}(A) = G^{-}(A')$, P-a.s.. For simplicity, in what follows we write A instead of $[A]$ and speak of sets instead of equivalence classes.

Core-determining classes, critical sets, and implicit-equality sets are defined as follows.

Definition 5.1 (Core-Determining Class). *For a class of sets $\mathcal{C} \subseteq \mathfrak{F}$, denote $\mathcal{M}_Q(\mathcal{C}) = \{\mu \ll Q : \mu(A) \geqslant C_G(A) \text{ for all } A \in \mathcal{C}\}$. A class $\mathcal{C} \subseteq \mathfrak{F}$ is core-determining if $\mathcal{M}_Q(\mathcal{C}) = \mathcal{M}_Q(\mathfrak{F})$.*

Definition 5.2 (Critical and Implict-Equality Sets). *A set A is critical if $\mathcal{M}_Q(\mathfrak{F} \backslash \{A\}) \neq \mathcal{M}_Q(\mathfrak{F})$. A set A is an implicit equality set if $\mu(A) = C_G(A)$ for any $\mu \in Core(G) \cap \mathcal{M}_Q$.*

Any CDC must contain all critical sets and ensure that all implicit equality restrictions hold. Connected random sets are defined as follows.

Definition 5.3 (Connected Random Sets). *The random set G is connected if $P(N(A)) > 0$ for any $A \in \mathcal{B}$ with $Q(A) > 0$.*

The idea is that 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.

The notions of self- and complement-connected sets extend as follows.

Definition 5.4 (Self- and Complement-Connected Sets). *Let $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a connected random set in the sense of Definition 5.3. Then*

1. *A subset $A \subseteq \mathcal{Y}$ 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..*

2. *A subset $A \subseteq \mathcal{Y}$ 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..*

5.2 The Smallest Core-Determining Class with Infinite Support

The results below 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 connected random set. A subset $A \subseteq \mathcal{Y}$ is critical if and only if it is both self- and complement-connected.*

Lemma 4 (Implicit equality Sets). *Let Assumption 5.1 hold and $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set. Let $\mathcal{Y} = \bigcup_{l \geqslant 1} \mathcal{Y}_l$ denote the finest partition of $\mathcal{Y}$ 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 $i \neq j$. 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 Assumption 5.1 hold and $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set.*

1. *If G is connected, the class $\mathcal{C}^*$ of all critical sets is the smallest core-determining class.*

2. *If the outcome space $\mathcal{Y}$ can be partitioned as in Lemma 4, there are infinitely many core-determining classes, each of which is the smallest by inclusion. Specifically, letting $\mathcal{C}_l^*$ denote the class of all critical sets in $\mathcal{Y}_l$, characterized in Lemma 3, the class $\mathcal{C}_l^* = \bigcup_{j \geqslant 1} \mathcal{C}_j^* \cup \bigcup_{j \neq l} \mathcal{Y}_j$ is core-determining, for each $l \in \mathbb{N}$.*

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 implies that the sharp identified set for θ_0 is generally intractable. However, certain functionals of interest, $\phi(\theta_0) \in \mathbb{R}$, may still admit tractable 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.

5.3 Examples

The first example studies a model with interval-valued data. For related results, see Beresteanu, Molchanov, and Molinari (2012), Molinari (2020), and references therein.

Example 4 (Interval Data)**.** Let $Y^* \in \mathcal{Y}$ denote a continuous outcome variable and $X \in \mathcal{X}$ denote covariates. Suppose the researcher does not observe Y^* directly but has access to continuous random variables $Y_L, Y_U \in \mathcal{Y}$ such that $Y^* \in G(Y_L, Y_U) = [Y_L, Y_U]$. 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 primitive parameter of interest is the joint distribution $\theta_0 = P_{Y^*X}$.

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 for any $A \subseteq \mathcal{Y}$ with $Q(A) > 0$, $P(N(A)) > 0$, the random set G is 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 $P(Y^* \in A \,|\, X = x) \geqslant P([Y_L, Y_U] \subseteq A \,|\, X = x)$ 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 subvevtor, 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 the inequality

$$F_{Y^*|X=x}(b) - F_{Y^*|X=x}(a) \geqslant P(Y_L \geqslant a, Y_U \leqslant b|X = x) \tag{12}$$

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 (12) being violated at $a = q_{Y^*|X=x}(\tau_2)$, $b = q_{Y^*|X=x}(\tau_1)$. Instead, it is easy to verify 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, Gandhi, and Quint (2013), Chesher and Rosen (2017), and Molinari (2020).

Example 5 (Ascending Auctions)**.** Consider a symmetric 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. Let $F \in \mathcal{F}$ denote the joint distribution of ordered valuations $V = (V_{1:N}, \ldots, V_{N:N})$ supported on $S = \{v \in [0, \overline{v}]^N : v_1 \leqslant \cdots \leqslant v_N\}$, where the class $\mathcal{F}$ summarizes the assumptions on the information structure. The distribution F is assumed to be continuous. For simplicity, suppose there is no reserve price and minimal bid increment. The researcher observes the two largest bids $(B_{N-1:N}, B_{N:N})$[15] 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:

$$G(V; F) = [0, V_{N-1:N}] \times [V_{N-1:N}, V_N] \cap S.$$

[15] For example, if it is hard to link the bids to the bidders, one can only reliably use the top two bids.

As long as F is supported on S, the support of $G(V;F)$ does not depend on F. Thus, the CDC is the same across F, and if the researcher has access to exogenous covariates Z independent of V, the Artstein's inequalities can be intersected over the values of Z.

It is easy to verify that the random set $G(V;F)$ is connected in the sense of Definition 5.3, so in view of 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.

However, the joint distribution F is typically of interest only to the extent that it allows us to calculate some counterfactual quantities. Aradillas-López, Gandhi, and Quint (2013) note that in ascending auctions, 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, Gandhi, and Quint (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.

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 $A = S \cap [a, \bar{v}] \times [0, b]$ for $b > a$ corresponds to the Artstein's inequality

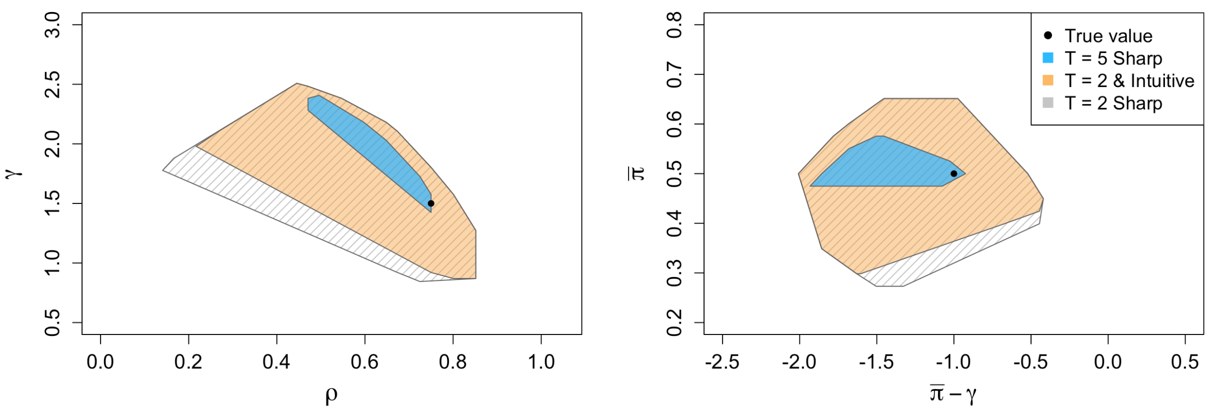


Figure 5: Projections of identified sets in the dynamic entry model from Example 2.

Note: The figure depicts convex hulls of the projections.

$F_{N-1:N}(b) - F_{N-1:N}(a) \geqslant P(B_{N-1:N} \geqslant a, B_{N:N} \leqslant b)$, 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.

6.1 A Dynamic Entry Model

In the first simulation exercise, we revisit the dynamic entry model of Berry and Compiani (2020) and 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. Specifically, they use 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 $T = 5$ obtained using 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 B. 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

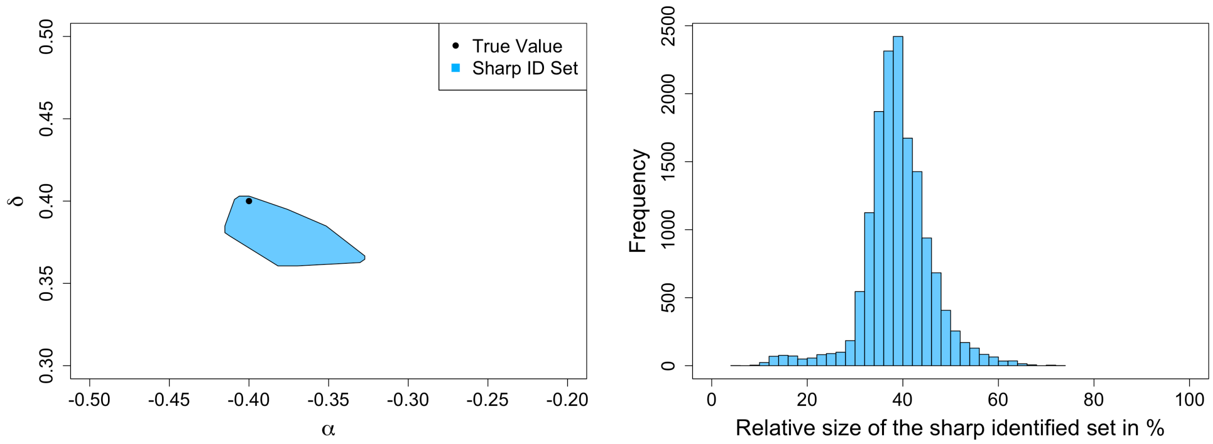


Figure 6: Size of the sharp identified set relative to identified sets constructed with the same number of inequalities in a market entry model with complementarities in Example 1.

intuitive inequalities of Berry and Compiani (2020); 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.

6.2 A Static Entry Model

In the second simulation exercise, we aim to quantify how much identifying information would be lost if the researcher used different equally-sized collections of inequalities for the analysis 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 combinations of 14 inequalities out of 254 ($\approx 10^{22}$ options), which is computationally infeasible. To approximate such an experiment, we sample 14 out of 254 inequalities at random 15,000 times and compute the corresponding identified sets using a fixed grid of parameter values. For each sample, we compute the relative size of the sharp identified set to the simulated one as the ratio of the numbers of grid points that satisfy the respective inequalities. The distribution of the relative sizes across simulations illustrates how alternative collections of inequalities perform relative to the smallest CDC.

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 the outcome space $\{(1,0,0),(0,1,0),(0,0,1),(1,1,1),(0,0,0),(1,1,0),(1,0,1),(0,1,1)\}$ 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.

7 Conclusion

In a large class of partially identified models, the sharp identified sets can be characterized by the so-called Artstein's inequalities, many of which may be redundant. To guide inequality selection, the literature has focused on finding core-determining classes — i.e., subsets of the inequalities that suffice for extracting all of the identifying information from the data and maintained assumptions. In this paper, we derived the smallest possible core-determining class, 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 can be applied far beyond the examples considered in the paper. Determining which moment inequalities are more informative for inference in finite samples is an important open question and natural direction for future research.

References

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

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

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

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

Bai, Y., S. Huang, S. Moon, A. M. Shaikh, and E. J. Vytlacil (2024): "On the Identifying Power of Monotonicity for Average Treatment Effects," .

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

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

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

——— (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79(6), 1785–1821.

——— (2012): "Partial identification using random set theory," *Journal of Econometrics*, 166(1), 17–32.

Berry, S., and E. Tamer (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*, pp. 889–917.

Berry, S. T., and G. Compiani (2020): "An instrumental variable approach to dynamic models," Discussion paper, National Bureau of Economic Research.

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

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

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

——— (2020): "Generalized instrumental variable models, methods, and applications," in *Handbook of Econometrics*, vol. 7, pp. 1–110. Elsevier.

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

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

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

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

Gu, J., T. Russell, and T. Stringham (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 E. Tamer (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111(1), 1–51.

Heckman, J. J., and R. Pinto (2018): "Unordered monotonicity," *Econometrica*, 86(1), 1–35.

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

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

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

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

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

Lee, S., and B. Salanié (2018): "Identifying effects of multivalued treatments," *Econometrica*, 86(6), 1939–1963.

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

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

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

Manski, C. F., and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

——— (2009): "More on monotone instrumental variables," *The Econometrics Journal*, 12, S200–S216.

Manski, C. F., and C. Sims (1994): "The selection problem," in *Advances in Econometrics, Sixth World Congress*, vol. 1, pp. 143–70.

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 F. Molinari (2018): *Random sets in econometrics*, vol. 60. Cambridge University Press.

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

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

Mourifie, I., M. Henry, and R. Meango (2020): "Sharp bounds and testability of a Roy model of STEM major choices," *Journal of Political Economy*, 128(8), 3220–3283.

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

——— (2013): "Max flows in O (nm) time, or better," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 765–774.

Pakes, A., and J. Porter (2024): "Moment inequalities for multinomial choice with fixed effects," *Quantitative Economics*, 15(1), 1–25.

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

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

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.

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., A. Torgovitsky, and H. Yang (2019): "Nonparametric estimates of demand in the california health insurance exchange," Discussion paper, National Bureau of Economic Research.

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

A Proofs from the Main Text

A.1 Lemma 1

The "only if" direction follows immediately from the arguments in Section 2.3: if a set A is such that $(A, G^{-}(A))$ is disconnected, the first argument applies; if $(A^c, G^{-1}(A^c))$ is disconnected, the second argument applies. For the "If" direction, we show that for any set A that is both self- and complement-connected, there exists a probability measure $\mu \in \text{Core}(G)$ satisfying $\mu(A) = C_G(A)$ and $\mu(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$. This implies that such A represents the fact of the convex polytope characterizing the $\text{Core}(G)$ and thus it is critical.

Say that a set A 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 $\nu \in \mathcal{M}$ be any probability distribution with $\nu(y) > 0$ for all $y \in \mathcal{Y}$. Define a Markov kernel $\pi_0 : \mathcal{U} \times 2^{\mathcal{Y}} \to [0, 1]$ as $\pi_0(u, A) = \frac{\nu(A \cap G(u))}{\nu(G(u))}$. Notice that $\pi_0(u, \cdot)$ is a probability measure supported on $G(u)$ with $\pi_0(u, A) > 0$ if and only if $u \in G^{-1}(A)$. Such π_0 induces a probability distribution $\mu_0 \in \mathcal{M}$ given by

$$\begin{aligned}
\mu_0(A) &= \textstyle\sum_{u \in \mathcal{U}} \pi_0(u; A) P(u) \\
&= \textstyle\sum_{u \in G^{-}(A)} \pi_0(u; A) P(u) + \sum_{u \in N(A)} \pi_0(u; A) P(u) \\
&= C_G(A) + \textstyle\sum_{u \in N(A)} \pi_0(u; A) P(u),
\end{aligned}$$

where $N(A) = G^{-1}(A) \backslash G^{-}(A)$. Consider a set A that is both self-and complement-connected. Define a Markov kernel π as

$$\pi(u, B) = \begin{cases} \frac{\pi_0(u, B \cap A^c)}{1 - \pi_0(u, A)} & u \in N(A) \\ \pi_0(u, B) & u \notin N(A). \end{cases}$$

Such π moves probability mass away from A so that the induced distribution $\mu \in \mathcal{M}$ satisfies $\mu(A) = C_G(A)$ by construction. In turn, for any set $\tilde{A} \neq A$ with $C_G(\tilde{A}) > 0$,

$$\mu(\tilde{A}) = C_G(\tilde{A}) + \sum_{u \in N(\tilde{A}) \cap N(A)} \frac{\pi_0(u, \tilde{A} \cap A^c)}{1 - \pi_0(u, A)} P(u) + \sum_{u \in N(\tilde{A}) \cap N(A)^c} \pi_0(u, \tilde{A}) P(u). \tag{A.1}$$

If $N(\tilde{A}) \cap N(A)^c \neq \varnothing$, the second sum in (A.1) is strictly positive and the desired result follows. It remains to consider the case $N(\tilde{A}) \subseteq N(A)$. There are three possibilities:

1. $A \cap \tilde{A} \neq \varnothing$ and $A \cap \tilde{A}^c \neq \varnothing$. Since $N(\tilde{A}) \subseteq N(A)$, in particular, $N(\tilde{A}) \cap G^{-}(A) = \varnothing$. In this case, the sets $A_1 = A \cap \tilde{A}$ and $A_2 = A \cap \tilde{A}^c$ satisfy $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \varnothing$ and

$G^{-}(A) = G^{-}(A_1) \cup G^{-}(A_2)$, which contradicts the assumption that A is self-connected.

2. $A \cap \tilde{A} = \varnothing$. In this case, the first sum in (A.1) is strictly positive.

3. $A \cap \tilde{A}^c = \varnothing$. Then, there cannot exist $u \in \mathcal{U}$ such that $G(u) \cap (\tilde{A} \cap A^c) \neq \varnothing$ and $G(u) \cap \tilde{A}^c \neq \varnothing$. In this case, the sets $A_1^c = \tilde{A} \cap A^c$ and $A_2^c = \tilde{A}^c$ satisfy $A_1^c \cup A_2^c = A^c$, $A_1^c \cap A_2^c = \varnothing$ and $G^{-1}(A^c) = G^{-1}(A_1^c) \cup G^{-1}(A_2^c)$, which contradicts the assumption that A is complement-connected.

Therefore, there exists a distribution $\mu \in \mathcal{M}$ such that $\mu(A) = C_G(A)$ and $\mu(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$, which means that A is critical.

A.2 Lemma 2

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 $\mu(\mathcal{Y}_l) = P(U \in G^{-}(\mathcal{Y}_l)) = C_G(\mathcal{Y}_l)$. To see that no other subset $A \subseteq \mathcal{Y}$ satisfies this property, consider the Markov kernel π_0 and the induced distribution μ_0 from the proof of Lemma 1. Since $N(A) \neq \varnothing$, it follows that $\mu_0(A) > C_G(A)$, so such A cannot be an implicit equality set.

A.3 Theorem 1

First, suppose $\mathbf{B}$ is connected, so there are no implicit-equality sets. To prove that the class $\mathcal{C}^*$ of all critical sets is core-determining, we show that all non-critical sets can be removed simultaneously without changing the core. Doing so would only be problematic if there existed distinct non-critical sets A and B such that A must be present to claim that B is redundant and vice versa (i.e., no collection of sets excluding A can suffice to claim B is redundant, and vice versa). By Lemma 1, every non-critical set must be not self-connected or not complement-connected. Consider three possible cases.

1. Both A and B are not self-connected. Then, $A = \tilde{A} \cup B$ for some $\tilde{A}$ with $\tilde{A} \cap B = \varnothing$ and $G^{-}(A) = G^{-}(\tilde{A}) \cup G^{-}(B)$, and also $B = \tilde{B} \cup A$ for some $\tilde{B}$ with $\tilde{B} \cap A = \varnothing$ and $G^{-}(B) = G^{-}(\tilde{B}) \cup G^{-}(A)$. This implies $A = B$.

2. Both A and B are not complement-connected. Then, $A^c = \tilde{A}^c \cup B^c$ for some $\tilde{A}$ with $\tilde{A}^c \cap B^c = \varnothing$ and $G^{-1}(\tilde{A}^c) \cup G^{-1}(B) = \varnothing$, and also $B^c = \tilde{B}^c \cup A^c$ for some $\tilde{B}$ with $\tilde{B}^c \cap A^c = \varnothing$ and $G^{-1}(\tilde{B}^c) \cap G^{-1}(A^c) = \varnothing$. This implies $A = B$.

3. A is not self-connected and B is not complement-connected. Then, (i) $A = \tilde{A} \cup B$ for some $\tilde{A}$ with $\tilde{A} \cap B = \varnothing$, and $G^{-}(A) = G^{-}(\tilde{A}) \cup G^{-}(B)$, and also (ii) $B^c = \tilde{B}^c \cup A^c$

for some $\tilde{B}$ with $\tilde{B}^c \cap A^c = \varnothing$ and $G^{-1}(\tilde{B}^c) \cap G^{-1}(A^c) = \varnothing$. Then, by construction $\tilde{B}^c = \tilde{A}$, which implies that the random set G cannot be connected. Indeed, for any $u \in \mathcal{U}$ such that $G(u) \cap \tilde{A} \neq \varnothing$ and $G(u) \cap B \neq \varnothing$, it must also be that $G(u) \cap A^c \neq \varnothing$, because otherwise (i) is violated—but the existence of such u would contradict (ii).

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 $G^{-1}(\mathcal{Y}_i) \cap G^{-1}(\mathcal{Y}_j) = 0$. 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 $(A_l)_{l=1}^{L}$. 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. Combined with the preceding argument, these facts imply that a class containing all critical subsets of all $\mathcal{Y}_l$ and all but one implicit equality sets is one of the smallest CDCs.

A.4 Algorithms 2 and 3

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.

A.5 Lemma 3

Under the stated assumptions, the proof is nearly identical to that of Lemma 1 with the following modifications. Let $\nu \ll Q$ be a probability measure that satisfies $d\nu/dQ > 0$, so that $\nu(G(u)) > 0$, P-a.s.. Define a map $\pi_0 : \mathcal{U} \times \mathcal{B} \to [0,1]$ as $\pi_0(u, A) = \frac{\nu(A \cap G(u))}{\nu(G(u))}$. By

Robbins' Theorem (see Theorem 1.5.16 in Molchanov, 2005) and standard properties of measurable functions, the map $u \mapsto \pi_0(u, A)$ is measurable for each $A \in \mathcal{B}$. By construction, the set-function $A \mapsto \pi_0(u, A)$ defines a probability measure on $\mathcal{B}$, for P-almost all $u \in \mathcal{U}$. Thus, π_0 is a Markov kernel. Notice that $\pi_0(u, \cdot)$ is supported on $G(u)$, and Definition 5.3 guarantees $\pi_0(u, A) > 0$ for P-almost all $u \in N(A)$. Such π_0 induces a probability distribution $\mu_0 \ll \nu$ given by

$$\begin{aligned}\mu_0(A) &= \int_{G^{-1}(A)} \pi_0(u; A) dP \\ &= \int_{G^{-}(A)} \pi_0(u; A) dP(u) + \int_{N(A)} \pi_0(u; A) P \\ &= C_G(A) + \int_{N(A)} \pi_0(u; A) dP.\end{aligned}$$

The rest of the argument proceeds exactly as in the proof of Lemma 1, with all summations replaced by integrals and qualifiers P-a.s. and Q-a.s. added when referring to set operations in $\mathcal{U}$ and $\mathcal{Y}$.

A.6 Lemma 4 and Theorem 2

The proofs are nearly identical to the corresponding proofs of Lemma 2 and Theorem 1, with the qualifiers P-a.s. and Q-a.s. added when referring to set operations in $\mathcal{U}$ and $\mathcal{Y}$.

B Simulation Design in Section 6

B.1 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 $X_1 \sim \text{Bernoulli}(p = 0.5)$ 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 $\rho \in [0, 1]$.

C 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, Rosen, and Smolinski (2013) and Tebaldi, Torgovitsky, and Yang (2019).

Example 6 (Discrete Choice with Endogeneity)**.** Individuals choose one of $J+1$ alternatives, $Y \in \{y_0, y_1, \ldots, y_J\} \equiv \mathcal{Y}$, where y_0 represents the outside option. Choosing y_j yields utility $v_j(X) + \varepsilon_j$, where $X \in \{x_1, \ldots, x_K\} \equiv \mathcal{X}$ 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 $v_0 = 0$ and $\varepsilon_0 = 0$. Some components of X may be correlated with the latent payoff shifters $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_J)$, but the nature of this dependence is left unspecified. The econometrician observes $Y \in \mathcal{Y}$ and $X \in \mathcal{X}$, and has access to 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 $U = (U_1, \ldots, U_J) \in \mathbb{R}^J$. Then, given U and θ, the model produces a set of possible values for (Y, X) given by

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

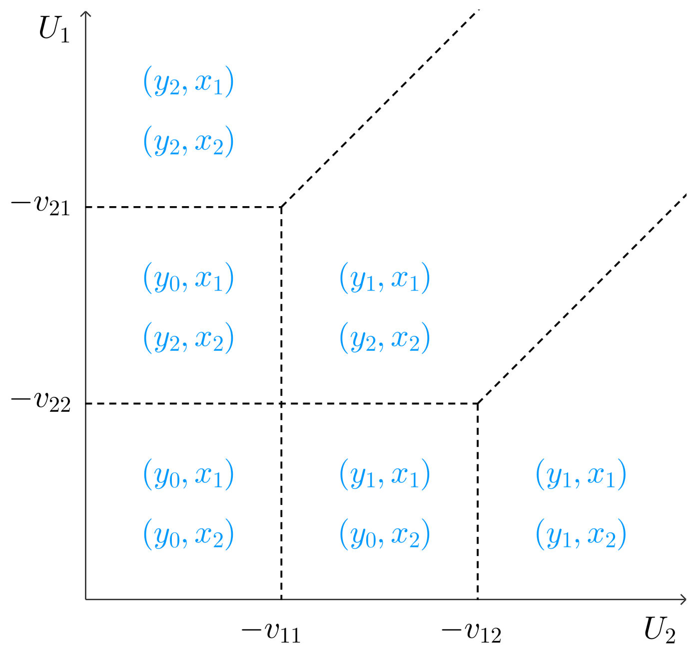
Figure 7 illustrates possible realizations of $G(U; \theta)$ for some fixed θ in a model with $\mathcal{Y} = \{y_0, y_1, y_2\}$ and $X \in \{x_1, x_2\}$. Dashed lines outline the partition of the latent variable space that corresponds to possible realizations of $G(U; \theta)$, highlighted in blue.

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.[16] 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

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

If $X \in \{x_1, \ldots, x_K\}$, the power set of the outcome space grows proportionally to $2^{(J+1)K}$.

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



with $J = K = 2$.

Figure 7: Set-valued predictions in a discrete choice model from Example 6

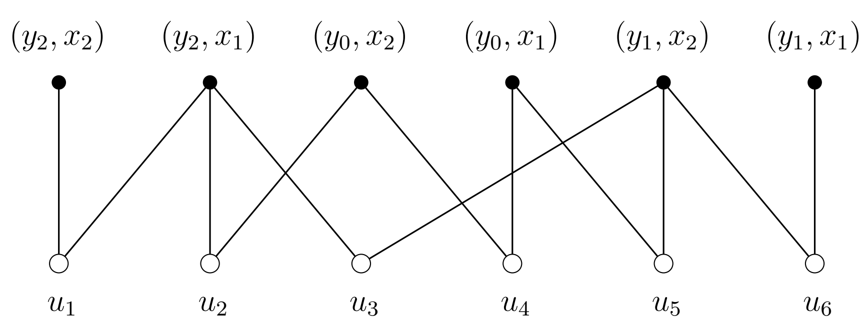


Figure 8: Discrete choice model from Example 6 with $J = 2$ and $X \in \{x_1, x_2\}$.

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 3: Core-determining classes in the discrete choice model from Example 6.

Yet, due to the simple structure of the underlying bipartite graph, the smallest CDC appears to grow proportionally to 2^K. Table 3 summarizes the results for $K \in \{2, \ldots, 15\}$.

The analysis above is similar to Chesher, Rosen, and Smolinski (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, Rosen, and Smolinski (2013), yet we also show that the resulting characterization cannot be further simplified, providing a somewhat negative result. Tebaldi, Torgovitsky, and Yang (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, $\eta = (\eta_1, \ldots, \eta_{|MRP|})$, 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 7 (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. ■