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: December 30, 2024

Abstract

One of the main challenges in partially-identified models is obtaining a tractable characterization for the sharp identified set, which exhausts all information contained in the data and modeling assumptions. In a large class of models, sharp identified sets can be described using a special kind of moment inequalities known as Artstein's inequalities. While theoretically convenient, this approach is often considered impractical because the total number of inequalities is too large. However, many of the inequalities may be redundant in the sense that excluding them does not lose identifying information. In this paper, we characterize the smallest possible collection of inequalities that describes the sharp identified set and provide an efficient algorithm for finding such inequalities in practice. In situations when the smallest set of inequalities is infeasible, we discuss imposing additional assumptions that facilitate inequality selection without losing sharpness. We apply the results to the models of static and dynamic market entry, discrete choice, potential outcomes models, selectively-observed data, and English auctions and conduct a simulation study to demonstrate that the proposed method substantially improves upon ad hoc 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 comments and discussions that helped improve this paper. 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 "true" parameter value θ_0. 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), 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 with heterogeneous or counterfactual choice sets (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 inequality

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

must hold for all $A \subseteq \mathcal{Y}$, all $x \in \mathcal{X}$. 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 $Y \in G(U, X; \theta_0)$, almost surely. Thus, assuming the parameter space Θ captures all other assumptions imposed on the model, the identified set Θ_0 is sharp.

The characterization in (2) is theoretically appealing but is often impractical since the total number of inequalities is very large. In such cases, it is customary to select a smaller collection of inequalities based on intuition and proceed with an outer set for Θ_0. This approach has two important drawbacks. On the one hand, the implied bounds on the

[1] In some of the models cited below, the set-valued predictions naturally arise in the space of latent variables. That is, 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 "true" parameter value θ_0. The approach proposed in the paper applies symmetrically in such settings.

parameters of interest may be very wide due to a loss of identifying information. On the other hand, having non-sharp identified sets that are very narrow may be a signal of "identification by misspecification" and lead to misleading conclusions (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 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 losses and 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 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, such $\mathcal{C}$ leads to a more concise characterization of the sharp identified set. We propose a simple analytical criterion to determine whether a particular inequality is redundant and use it to derive the smallest possible CDC. We show that the CDC depends only on the structure of the model's correspondence G and the null sets of the underlying probability distribution, so in discrete-outcome models, it typically needs to be computed only a finite number of times. Additionally, we 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, representing the model's correspondence, and its computational complexity is proportional to the size of the smallest CDC. In settings where even the smallest CDC is infeasible, we discuss imposing additional assumptions to motivate inequality selection without losing sharpness.

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 $\Theta_0 = \{\theta \in \Theta : P_{Y|X=x} \in \mathcal{P}(x;\theta),\ x \in \mathcal{X}\text{-a.s.}\}$. The 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, the authors represent the set $\mathcal{P}(x;\theta)$ as the core of a distribution of a random set $G(U, x;\theta)$, conditional on $X = x$, characterized by 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, discussed in more detail in Section 4.4, 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. More generally, even the smallest CDC may grow exponentially with the size of the outcome space and is much harder to characterize. This paper extends the results of Galichon and Henry (2011) by relating the smallest CDC to the structure of the model's correspondence more generally and proposes an efficient algorithm to compute it in practice. Chesher and Rosen (2017) derive analytical *sufficient* conditions for identifying redundant Artstein's inequalities. By adding an extra condition, we obtain a set of *necessary and sufficient* conditions for redundancy and use it to characterize the smallest possible CDC.

Luo and Wang (2018) (LW18) also provide a characterization of the smallest CDC, which they call "exact," in their Theorem 2. We improve and extend this result in several directions. First, although Theorem 1 below delivers the same CDC as Theorem 2 of LW18, coupled with Lemmas 1 and 2, it provides a more transparent and complete characterization, including the proofs. These results clearly identify the "critical" sets, which must be included in any CDC, as well as "implicit-equality" sets, for which Artstein's inequalities hold as equalities, providing useful information for inference. Second, Corollary 1.1 establishes that the smallest CDC depends only on the support of a random set. 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, allowing for a simpler characterization in many settings. Third, Theorem 1 implies an efficient algorithm for computing the smallest CDC numerically, which remains feasible far beyond Algorithm 1 of LW18. Finally, Section 5 extends the main results to the settings with continuous outcomes.

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. They characterize the Aumann expectation via the support function, thus expressing the sharp identified set through a convex optimization problem. 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 defining their facets can be computed by solving a multiple-objective linear program. We argue that the CDC approach is complementary to these methods, leading to faster computation of the identified set and simpler inference procedures whenever the smallest CDC is manageable.

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 in 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 existing methods. Section 5 provides an extension to models in which the outcomes have infinite support. Section 6 illustrates the utility of selecting inequalities. Section 7 concludes.

2 Models with Set-Valued Predictions

2.1 Motivating Examples

To outline the scope of the paper, we start with four stylized examples, all of which deal with discrete-outcome models. Two additional examples are considered in Section 3.3, and a discussion of continuous-outcome models is deferred to Section 5.

The first example is a static entry game studied in 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:

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

where $Y = (Y_1, \ldots, Y_N)$ is the outcome vector, $N_j(Y)$ is number of entrants except j, $U = (\varepsilon_1, \ldots, \varepsilon_N)$ are idiosyncratic payoff shifters, and $(\alpha_j, \delta_j)_{j=1}^N$ are payoff parameters. The joint distribution of latent variables U, denoted $F(\cdot; \gamma)$, is assumed known up to a

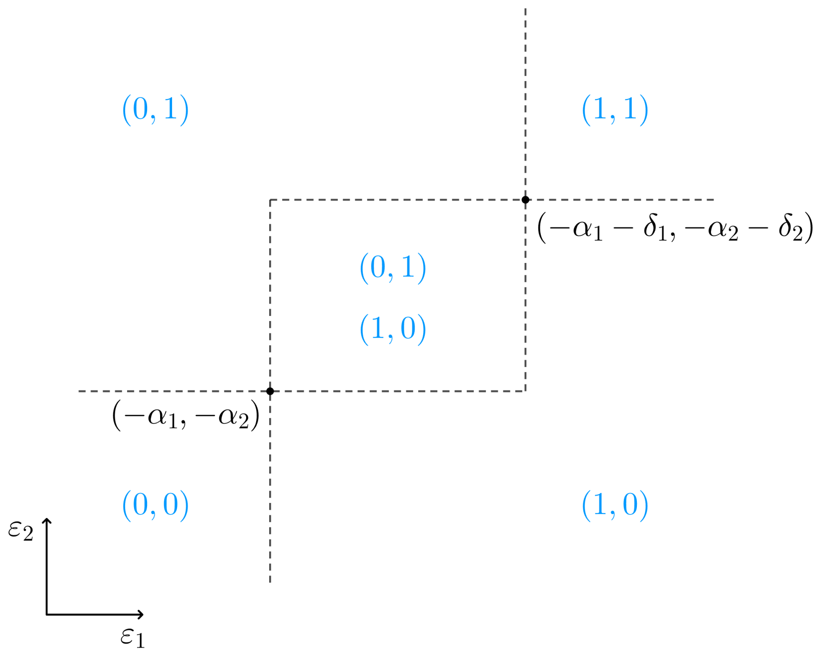


Figure 1: Set-valued predictions in the entry game from Example 1 with $N = 2$ and $\delta_j < 0$ for $j = 1, 2$.

final dimensional parameter γ. Exogenous covariates 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 $Y \in (Y_1, \ldots, Y_N) \in \{0, 1\}^N$.

Given U and θ, the model produces a set of predictions for the outcomes Y corresponding to the set of pure-strategy Nash Equilibria:

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

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

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

Example 2 (Dynamic Monopoly Entry)**.** In time period $t = 1, \ldots, T$, a firm decides to stay

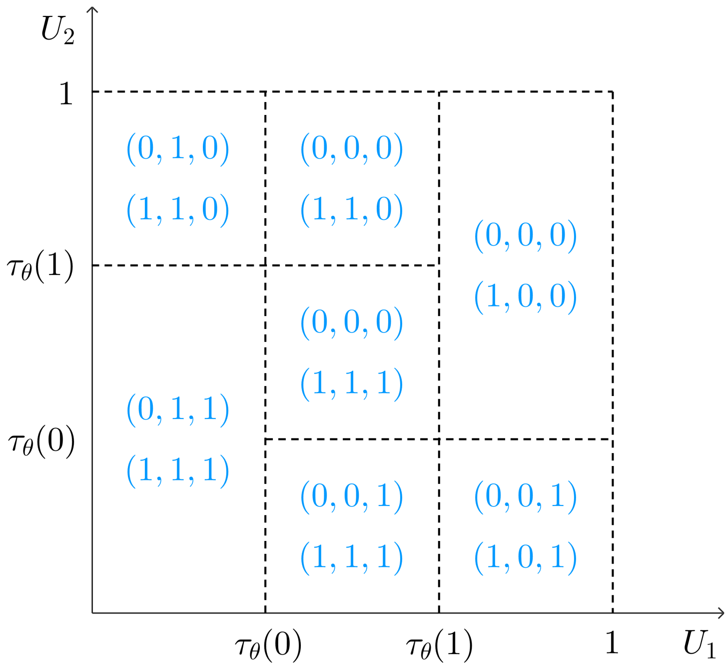


Figure 2: Set-valued predictions in the dynamic model from Example 2 with $T = 2$. Outcomes are labeled as (X_1, A_1, A_2).

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 if 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 fixed profit and sunk costs of entering the market correspondingly. 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 reserarcher observes $Y = (X_1, A_1, \ldots, A_T) \in \{0, 1\}^{T+1}$.

The Bellman equation for the firm's problem is:

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

Under standard conditions, there is a unique stationary solution:

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

where U_t is the quantile transformation of ε_t, and τ is an increasing function of X_t known up to 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 $Y = (X_1, A_1, \ldots, A_T)$. Then, given $U = (U_1, \ldots, U_T)$ and θ, the model produces a set of possible values for Y given by:

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

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

The next example is a discrete-choice model with endogenous covariates, studied in Chesher, Rosen, and Smolinski (2013) and Tebaldi, Torgovitsky, and Yang (2019).

Example 3 (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 $v_j(X) + \varepsilon_j$ (with $v_0 = 0$ and $\varepsilon_0 = 0$), where $X \in \{x_1, \ldots, x_k\} \equiv \mathcal{X}$ may include prices, individual-level, and market-level covariates, and $\varepsilon_j \in \mathbb{R}$ are latent utility shifters. Individuals maximize their utility, so $Y = y_{j^*}$ for $j^* = \operatorname{argmax}_j\{v_j(X) + \varepsilon_j\}$. 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 $Z \in \mathcal{Z}$ statistically independent of ε.

Note that X is endogenous and its data-generating process is left unspecified. As in the previous example, 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 3 illustrates possible realizations of $G(U;\theta)$ for some fixed θ in a model with $J = K = 2$. The dashed lines outline the partition of the latent variable space corresponding to the possible realizations of $G(U;\theta)$, highlighted in blue. ■

The final example is a potential outcomes model, studied, in particular, in 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), Mouri-

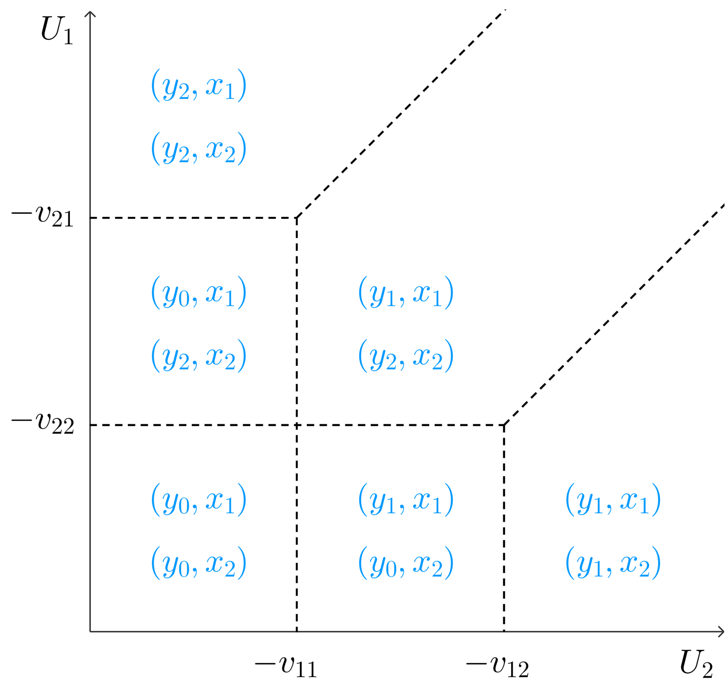


Figure 3: Set-valued predictions in a discrete choice model from Example 3 with $J = K = 2$.

fie, Henry, and Meango (2020), Russell (2021), and Bai, Huang, Moon, Shaikh, and Vytlacil (2024), among many others.

Example 4 (Potential Outcomes Models)**.** Let $\mathcal{D}$, $\mathcal{Y}$, and $\mathcal{Z}$ be some finite sets. Let $D \in \mathcal{D}$ denote a 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 desired restrictions on the outcome response function $d \mapsto Y_d^*$ (e.g., monotonicity) are summarized by $Y^* \in \mathcal{S}_{Y^*}$ for some known set $\mathcal{S}_{Y^*} \subseteq \mathcal{Y}^{|\mathcal{D}|}$. The primitive parameter of interest is the joint distribution of potential outcomes, $\theta = \{P(Y^* = y^*)\}_{y^* \in \mathcal{Y}^{|\mathcal{D}|}}$.

Given (Y, D, Z), the model produces a set of possible values for the latent vector of potential outcomes Y^*. Specifically, if $D = d$, then $Y_d^* = Y$, but the only information available about $Y_{\tilde{d}}^*$ for $\tilde{d} \neq d$ is that $Y_{\tilde{d}} \in \mathcal{Y}$, and $Y^* \in \mathcal{S}_{Y^*}$. Thus, the model produces a set-valued prediction for Y^* is given by:

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

where $B_d(Y) = (\mathcal{Y} \times \cdots \times \{Y\} \times \ldots \mathcal{Y})$ with $\{Y\}$ in the d-th coordinate. Notice that Z

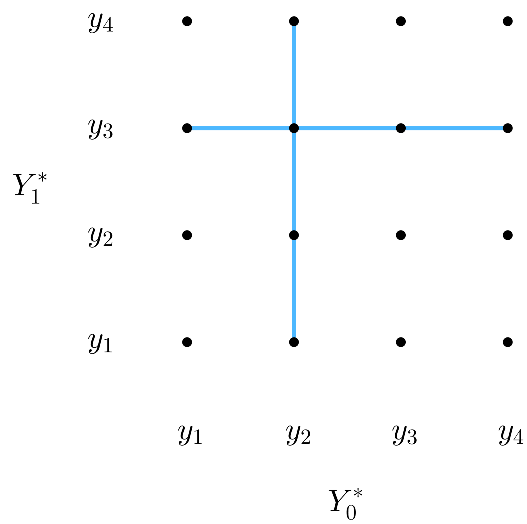


Figure 4: Set-valued predictions in a potential outcomes model from Example 4 with $|\mathcal{D}| = 2$ and $|\mathcal{Y}| = 4$.

does not affect $G(Y, D)$ in any way. Figure 4 illustrates possible realizations of $G(Y, D)$ for $|\mathcal{D}| = 2$ and $|\mathcal{Y}| = 4$, and $\mathcal{S}_{Y^*} = \mathcal{Y}^2$. The vertical blue line corresponds to $G(y_2, 0)$, and the horizontal blue line to $G(y_3, 1)$. ∎

2.2 Background: Random Sets and Artstein's Inequalities

In the above examples, the set of outcomes (or latent variables) predicted by the model depends on a realization of random variables, so it is a random set. Naturally, identified sets in such settings can be conveniently described using the language of random sets. We briefly introduce the necessary concepts below and refer the reader to Molchanov and Molinari (2018) for a textbook treatment.

Let $(\mathcal{U}, \mathcal{F}, P)$ be a probability space and $(\mathcal{Y}, \mathcal{A})$ a finite measurable space, with $\mathcal{Y} = \{y_1, \dots, y_S\}$ and $\mathcal{A} = 2^{\mathcal{Y}}$. Let $\mathcal{M}$ denote the set of all probability measures on $\mathcal{Y}$, and $\mathfrak{F}$ denote the class of all subsets of $\mathcal{Y}$. Let $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ be a correspondence. For each $A \subseteq \mathcal{Y}$, 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 a sense that $G^{-}(A) \in \mathcal{F}$ for all $A \in \mathfrak{F}$, it defines a *random (closed) set*. Its distribution can be described

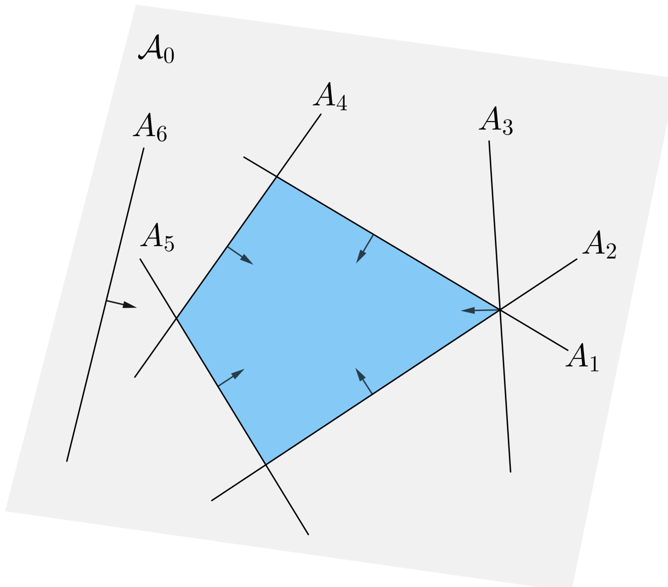


Figure 5: The core of a random set.

by the so-called *containment functional*, defined for all $A \in \mathfrak{F}$ as:

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

The *support* of a random set G, denoted $S \subseteq \mathfrak{F}$, is the collection of sets $A \in \mathfrak{F}$ such that $P(G = A) > 0$. Any random variable $Y : (\mathcal{U}, \mathcal{F}, P) \to (\mathcal{Y}, \mathcal{A})$ satisfying $P(Y \in G) = 1$ 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 dominating the containment functional (on all closed sets):

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

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 $\mathcal{C} \subseteq \mathfrak{F}$, denote*

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

Say that a class $\mathcal{C} \subseteq \mathfrak{F}$ is core-determining if $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathfrak{F})$.

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

Definition 2.2 (Critical and Implicit Equality Sets)**.** *A set $A \in \mathfrak{F}$ is critical if $\mathcal{M}(\mathfrak{F} \backslash A) \neq \mathcal{M}(\mathfrak{F})$. A set $A \in \mathfrak{F}$ 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 5 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 indicating the direction in which it is satisfied. The core is highlighted in blue. Any class of sets including $\mathcal{A}_0 \cup \{A_1, A_2, A_4, A_5\}$ 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 implications between Artstein's inequalities. For what triplets of sets $A_1, A_2, A \in \mathfrak{F}$, do the ienqualities $\mu(A_1) \geqslant C_G(A_1)$ 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 when the containment functional is additive, the answer is fairly straightforward.

First, suppose 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.

A special case occurs when the set A cannot be expressed as a union of elements of the support of G, i.e., $A \neq G(G^-(A))$. 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 $\mu(A_1) \geqslant C_G(A_1)$,

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

so A is redundant given A_1.

Second, suppose that for some $A \subseteq \mathcal{Y}$, there are sets $A_1, A_2 \subseteq \mathcal{Y}$ such that $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, 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 $\mu(A_2) \geqslant C_G(A_2)$,

$$1 + \mu(A) = \mu(A_1) + \mu(A_2) \geqslant C_G(A_1) + C_G(A_2) = 1 + C_G(A), \tag{7}$$

so A is redundant given A_1 and A_2. Note that 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 $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \varnothing$.

As it turns out, no other non-trivial implications between Artstein's inequalities exist. This fact can be used to characterize all critical and implicit-equality sets in a transparent way and obtain a simple core-determining class. A different representation of random sets is more suitable for that purpose, as discussed below.

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:[2]

$$\Theta_0 = \{\theta \in \Theta : P_{Y|X=x}(A) \geqslant C_{G(U,x;\theta)}(A), \text{ for all } A \in \mathcal{C}(x, \theta), \text{ a.s. } x \in \mathcal{X}\}, \tag{8}$$

where $\mathcal{C}(x, \theta) \subseteq \mathfrak{F}$ is a core-determining class for the random set $G(U, x; \theta)$ conditional on $X = x$. Evidently, smaller core determining classes $\mathcal{C}(x, \theta)$ lead to a more tractable characterization. In this section, we characterize the smallest possible core-determining class $\mathcal{C}^*(x; \theta)$, clarify how it depends on x and θ, and illustrate the results in several widely-used settings.

3.1 Random Sets as Bipartite Graphs

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 $P_{(x,\theta)}(u_k) = P_{U|X=x}(\{u : G(u, x; \theta) = G_k\})$. In this way, the random set $G(U, x; \theta)$, conditional on $X = x$, can be viewed as $G : (\mathcal{U}(x,\theta), 2^{\mathcal{U}(x,\theta)}, P_{(x,\theta)}) \rightrightarrows \mathcal{Y}$, defined on a discrete probability space, so it can be represented by an undirected bipartite graph $\mathbf{B}$ with vertices $V(\mathbf{B}) = (\mathcal{U}, \mathcal{Y})$ and edges $E(\mathbf{B}) = \{(u, y) \in \mathcal{U} \times \mathcal{Y} : y \in G(u)\}$.

Example 1 – 4 (Continued)**.** Figure 6 presents the bipartite graphs for Examples 1 – 4.

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, $u_1 = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 : \varepsilon_j <$

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

$-\alpha_j, j = 1, 2\}$, and $u_3 = \{(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2 : -\alpha_j \leqslant \varepsilon_j < -\alpha_j - \delta_j\}$. Additionally, 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 2. 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 $u_5 = \{(U_1, U_2) \in [0, 1]^2 : U_1 > \tau_\theta(1), U_2 > \tau_\theta(0)\}$. Additionally, 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 discrete choice model from Example 3 with $\mathcal{Y} = \{y_0, y_1, y_2\}$ and $X \in \{x_1, x_2\}$. 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 3. For example, $u_4 = \{(U_1, U_2) : U_1 \leqslant -v_{11}, U_2 \leqslant -v_{22}\}$. Additionally, for example, $G^-(\{(y_2, x_1), (y_0, x_1)\}) = \varnothing$ and $G^{-1}(\{(y_2, x_1), (y_0, x_1)\}) = \{u_1, u_2, u_3, u_4, u_5\}$.

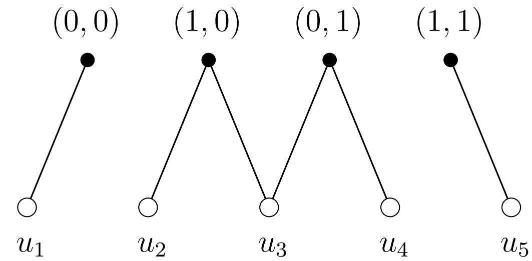
Panel (d) depicts the potential outcomes model from Example 4 with $\mathcal{D} = \{0, 1\}$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and $\mathcal{S}_{Y^*} = \mathcal{Y}^2$. The upper part is $\mathcal{S}_{Y^*}$, corresponding to the support of the latent potential outcome vector Y^*, and the lower part 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 Figure 4. Additionally, for example, $G^-(\{(2, 1), (2, 2), (2, 3), (2, 4)\}) = \{(0, 2)\}$, and $G^{-1}(\{(2, 1), (2, 2), (2, 3), (2, 4)\}) = \{(1, 1), (0, 2), (1, 2), (1, 3), (1, 4)\}$. ■

In practice, the bipartite graphs are easy to construct numerically; see Section 4.

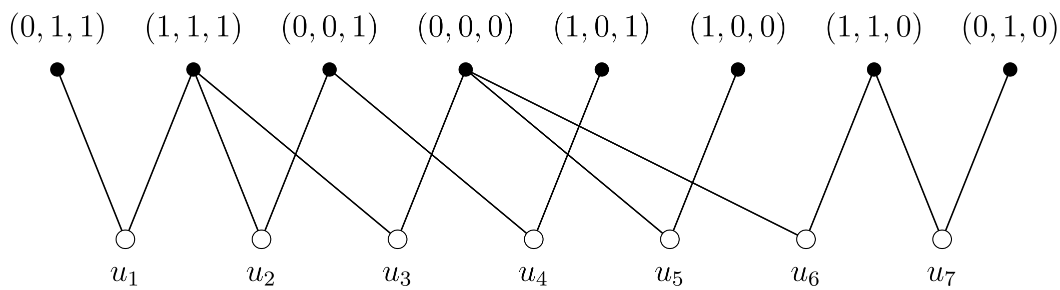
3.2 The Smallest Core-Determining Class

The implications between Artstein's inequalities discussed in Section 2.3 can be conveniently expressed in terms of 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.

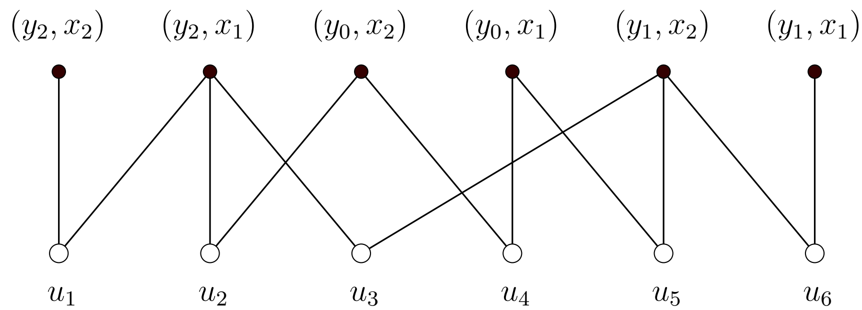
Consider the graph in Panel (b) of Figure 6. First, let $A_1 = \{(0, 1, 1), (1, 1, 1)\}$, $A_2 = \{(1, 1, 0), (0, 1, 0)\}$, and $A = A_1 \cup A_2$. Then, $G^-(A_1) = \{u_1\}, G^-(A_2) = \{u_7\}$, and $G^-(A) = \{u_1, u_7\}$. Thus, A is redundant given A_1 and A_2, as in (5). Note that in this case, the subgraph induced by $(A, G^-(A))$ is disconnected. Second, let $A = \{(0, 0, 1), (0, 0, 0), (1, 0, 1)\}$ 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 (6). In this case, again, the subgraph induced by $(A, G^-(A))$ is disconnected. Finally, let $A =$



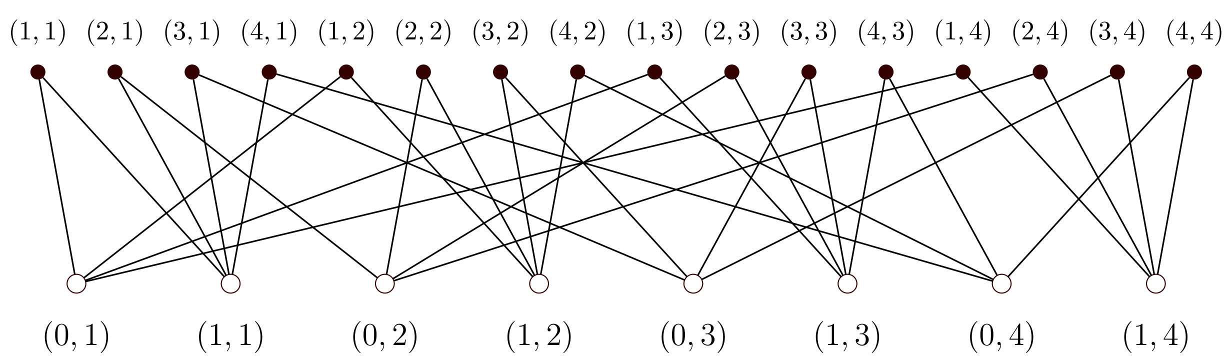
(a) Entry game from Example 1 with $N = 2$ and $\delta_j < 0$ for $j = 1, 2$.



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



(c) Discrete choice model from Example 3 with $J = 2$ and $X \in \{x_1, x_2\}$.



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

Figure 6: Bipartite graphs in Examples 1 – 4.

$\{(1,1,1),(0,0,1),(0,0,0)\}$, $A_1 = A \cup \{(0,1,1)\}$, $A_2 = A \cup \{(1,0,1),(1,0,0),(1,1,0),(0,1,0)\}$, 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 (7). Note that 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)–(7), the subgraph of $\mathbf{B}$ induced either by $(A, G^-(A))$ 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 by $(A^c, G^{-1}(A^c))$ are connected.*

The proof of this result is constructive: given a set A satisfying the above assumptions, we construct a distribution $\mu \in \mathrm{Core}(G)$ such that $\mu(A) = C_G(A)$ and $\mu(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$. This implies that A must be critical. The assumption $P(u_k) > 0$ for all $k = 1, \dots, K$ merely ensures 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 if $A = \mathcal{Y}_l$ for some $l = 1, \dots, L$.*

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 6, the implicit equality sets are $\{(0,0)\}$, $\{(1,1)\}$, and $\{(1,0),(0,1)\}$. In panels (b)–(d) there are no implicit equality sets. Combining the insights of Lemmas 1 and 2 yields a simple characterization of the smallest 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 $\mathcal{C}^* = \bigcup_{j=1}^{L} \mathcal{C}_l^* \cup \bigcup_{j \neq l} \mathcal{Y}_j$ is core-determining, for each $l = 1, \dots, L$.*

This result has two key implications. The first 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 $m \in \{1, \ldots, M\}$. Then, $\mathcal{C}^*(x,\theta) = \mathcal{C}^*_m(x)$ for all $\theta \in \Theta_m$, so the sharp identified set for θ can be expressed as:

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

Typically, it is also 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 given by:

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

Examples in Section 3.3 provide a detailed discussion.

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 the computational bottleneck of 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 $\mathbf{B}$. The details are discussed in Section 4.

For a fixed size of the outcome space, the size of the smallest CDC may be very different, depending on the structure of the correspondence G. In favorable scenarios, $|\mathcal{C}^*| \propto |\mathcal{Y}|$, and in the worst-case, $|\mathcal{C}^*| \propto 2^{|\mathcal{Y}|}$. Although it is generally impossible to estimate the size of the smallest CDC *ex ante*, some benchmarks can be provided when G has a special structure; see Appendix C.1.

3.3 Discussion and Applications

This section shows Theorem 1 in action, revisiting the four examples introduced in Section 2.1, and a directed network formation model from Gualdani (2021). The discussion points out how the smallest CDC depends on the unknown parameters of the model and whether it leads to a practical characterization of the sharp identified set.

In some settings, even the smallest CDC may be too large to be practically useful, so additional inequality selection is required. In this context, Examples 1 and 5 below discuss imposing additional restrictions on the structure of the model's correspondence or the equilibrium selection mechanism to simplify the analysis. The restrictions ensure that the resulting identified set Θ_0' is shaped by a subset of inequalities that characterize Θ_0, thus providing a formal basis for inequality selection: if the extra assumptions are satisfied, $\Theta_0' = \Theta_0$, but if they are violated, $\Theta_0' \supseteq \Theta_0$. In turn, Examples 2, 3, and 4 consider identification with instrumental variables.

Example 1 (Continued)**.** First, suppose $\delta_j < 0$ for all j, so the firms compete with each other upon entering the market. 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 6. 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 $j = 1, \dots, N$, a.s. $x \in \mathcal{X}$. Then, assuming additionally that $U = (\varepsilon_1, \dots, \varepsilon_N)$ and X are statistically independent, the sharp identified set for θ can be expressed as:

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

In this model, the set of Nash Equilibria can only contain equilibria with the same number of entrants, $n \in \{0, 1, \dots, N\}$, 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.[3] Table 1a summarizes the results for $N \in \{2, \dots, 6\}$. While 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 the firms 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 makes the corresponding bipartite graph very interconnected, which complicates identification. As before, while 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 $\delta_j > 0$, so, the CDC only needs to be computed once. Table 1a summarizes the

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

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.

The analysis can be simplified by restricting firm heterogeneity and profit functions. 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.[4] 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 $\beta_3 \geqslant \beta_2$. If $\beta_3 = \beta_2$, this is a direct simplification of the fully heterogeneous model discussed above. If $\beta_3 > \beta_2$, the firms compete in an asymmetric way (e.g., type-1 firms are large and type-2 firms are small).

From the identification perspective, a major simplification comes from the fact that the outcomes can now be aggregated into groups. For example, suppose there are three firms in total, one firm of the first type and two firms of the second type. Then, for any $(\varepsilon_1, \varepsilon_2)$, for which the outcome $(0, 1, 0)$ is an equilibrium, the outcome $(0, 0, 1)$ is also an equilibrium, by symmetry. Similarly, $(1, 1, 0)$ is an equilibrium whenever $(1, 0, 1)$ is. In the absence of restrictions on equilibrium selection, it is without loss of information to group such outcomes together and re-define the outcome space in terms of the numbers of entrants of each type. Then, letting N^t denote the number of potential entrants of type $t \in \{1, 2\}$, the new outcome space is $\tilde{\mathcal{Y}} = \{0, 1, \ldots, N^1\} \times \{0, 1, \ldots, N^2\}$. This leads to much simpler CDC-s. Table 1b shows that the smallest CDC remains tractable for different compositions of firm types. Extension to three or more types is straightforward.

Notably, the resulting set of inequalities is a subset of the inequalities in fully heterogeneous model. Therefore, the resulting sharp identified set Θ_0' always contains the sharp identified set Θ_0 in the fully heterogeneous model. If the type-heterogeneity assumption is

[4] 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 studied also in Beresteanu, Molchanov, and Molinari (2008), Galichon and Henry (2011) and Luo and Wang (2018).

N	2	3	4	5	6
Total	14	254	65,534	10^9	10^{19}
Smallest; $\delta_j < 0$	5	16	95	2,110	10^6
Smallest; $\delta_j > 0$	5	14	23,770	–	–

(a) Fully heterogeneous entry game

(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

(b) Entry game with two types and negative spillovers

Table 1: Core-determining classes in Example 1.

Note: Symbol "–" indicates infeasible to compute.

satisfied by the data, the two sets will coincide. In this way, the type-heterogeneity assumption suggests a formal basis for inequality selection. ■

Example 2 (Continued)**.** For $T = 2$, the relevant partition of the latent variable space is given in Figure 2, and the corresponding bipartite graph in Panel (b) of Figure 6. Note that 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 can be expressed as:

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

In this example, the bipartite graph $\mathbf{B}$ corresponding to the model's correspondence has a simple structure: each vertex u_j has exactly two neighbors, corresponding 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 2 summarizes the results for $T \in \{1, \ldots, 10\}$.

In more elaborated dynamic oligopoly models, discussed in 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 further research. ■

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

Table 2: Core-determining classes in the dynamic entry model from Example 2.

Example 3 (Continued)**.** Figure 3 depicts a possible partition of the latent variable space for $J = K = 2$, and Panel (c) of Figure 6 depicts the corresponding bipartite graph. Depending on the values of $\theta = (\{v_{jk}\}_{j,k}, \gamma)$, the partition and the probabilities of the corresponding regions are different, but provided $v_{11} > v_{12}$ and $v_{21} > v_{22}$, the corresponding bipartite graph remains the same. Suppose that all $\theta \in \Theta$ satisfy this restriction,[5] and let $Z \in \mathcal{Z}$ denote an excluded instrumental variable. 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:

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

If $X \in \{x_1, \ldots, x_K\}$, the power set of the outcome space grows proportionally to $2^{(J+1)K}$. Yet, due to the simple structure of the underlying bipartite graph, the smallest CDC appears to grow proportionally to 2^K. Table 3 summarizes the results for $K \in \{2, \ldots, 15\}$.

The analysis above is similar to Chesher, Rosen, and Smolinski (2013). The authors work with a different random set but also treat X as part of the outcome vector and condition only on Z, leaving $F_{U|X=x}$ completely unspecified. The inequalities in $\mathcal{C}^*$ coincide with the ones reported in Chesher, Rosen, and Smolinski (2013), yet we additionally show that the resulting characterization cannot be further simplified.

Tebaldi, Torgovitsky, and Yang (2019) take a different approach. They introduce the Minimal Relevant Partition (MRP), which is conceptually the same as the partition in Figure 3 (in the absence of counterfactual values of X). Additionally, they condition on both X and Z and treat the probabilities $\eta = (\eta_1, \ldots, \eta_{|MRP|})$ that the distribution $F_{U|X=x}$ assigns to each of the regions in MRP as unknown parameters. Theorem 2.33 in Molchanov and Molinari (2018) implies that the two approaches are equivalent and deliver the same identified sets. If the functional of interest depends only on η, and Z is discrete, the approach of Tebaldi, Torgovitsky, and Yang (2019) brings 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. See Section 4.4 for a related discussion. ■

[5] Otherwise, partition the parameter space as in Example 1.

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 3.

$\lvert\mathcal{Y}\rvert$	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

(a) No restrictions on 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	–	–

(b) Monotone 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	–

(c) Monotone and concave outcome response.

Table 4: Core-determining classes in the potential outcomes model from Example 4.

Notes: Panels (b) and (c) only report the size of the smallest core-determining class. Symbol "–" indicates that Algorithm 3 implemented in Julia took more than one minute.

Example 4. (Continued) Recall that the parameter of interest is the joint distribution of potential outcomes, $\theta = P_{Y*}$, with a known support $\mathcal{S}$. The support of the random set $G(Y,D)$ does not depend on P^*_Y, so no partition of parameter space is required and the smallest CDC needs to be computed only once.

First, consider the model without any restrictions on the support of Y^*. The corresponding bipartite graph, such as the one in Panel (d) in Figure 6, is connected, so there are no implicit-equality sets. Moreover, due to the special structure of the correspondence $G(Y,D)$, all critical sets can be described analytically. Recall the discussion preceding Lemma 1. Unions of elements of the support of $G(Y,D)$ are "lattice-shaped" sets $A = B_1 \times B_2 \cdots \times B_{|\mathcal{D}|}$, where all B_d are non-empty subsets of $\mathcal{Y}$ (as in Figure 4 for $|\mathcal{D}| = 2$). 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. Since Y^* and Z are independent, the corresponding Artstein's inequalities take the form $P(Y^* \in A) \geqslant P(G(Y,D) \subseteq A \,|\, Z = z)$, for each $z \in \mathcal{Z}$. In turn, if $B_d \subset \mathcal{Y}$ and $B_{\tilde{d}} = \mathcal{Y}$ for all $d \neq \tilde{d}$, the corresponding Artstein's inequalities restrict the marginal distribution of the Y^*_d. In this case, e.g., the inequality:

$$P(Y^* \in B_1 \times \mathcal{Y}^{|\mathcal{D}|-1}) \geqslant P(G(Y,D) \subseteq B_1 \times \mathcal{Y}^{|\mathcal{D}|-1} \,|\, Z = z)$$

is equivalent to

$$P(Y^*_1 \in B_1) \geqslant P(Y \in B_1, D = 1 \,|\, Z = z).$$

Notice the latter is implied by $P(Y^*_1 = y) \geqslant P(Y = y, D = 1 \,|\, Z = z)$ for all $y \in \mathcal{Y}$, so, in addition to the "lattice-shaped" sets described above, it suffices to consider sets of the form $A = \mathcal{Y} \times \cdots \times \{y\} \times \cdots \times \mathcal{Y}$ for each $y \in \mathcal{Y}$, for each $d \in \mathcal{D}$. Thus, the total number of critical sets, for each fixed $Z = z$, is:

$$\sum_{k=0}^{|\mathcal{D}|-2} \binom{|\mathcal{D}|}{k} (2^{|\mathcal{Y}|} - 2)^{|\mathcal{D}|-k} + |\mathcal{Y}||\mathcal{D}|.$$

The leading term in the above expression is of order $2^{|\mathcal{Y}||\mathcal{D}|}$, which is much smaller than the total number of inequalities, $2^{|\mathcal{Y}|^{|\mathcal{D}|}} - 2$. Panel (a) of Table 4 provides some examples.

Russell (2021) compares characterizations of the sharp identified set in this setting using three different approaches: (i) all Artstein's inequalities; (ii) the smallest available CDC; (iii) the dual approach of Galichon and Henry (2011). The author concludes that the CDC approach is never a preferred method as the results of Luo and Wang (2018) do not allow intersecting conditional Artstein's inequalities over the values of the instrument. However, since the support of the random set $G(Y,D)$ is not affected by the values of Z, Corollary 1.1

above implies that the smallest CDC does not change with Z, so the corresponding Artstein's inequalities can be intersected, so the sharp identified set is characterized by:

$$P(Y^* \in A) \geqslant \sup_{z\in\mathcal{Z}} P(G(Y,D) \in A \mid Z = z).$$

Thus, in this setting approach (ii) always yields a simpler characterization than (i). When the smallest CDC is very large and $\mathcal{Z}$ is relatively small, the dual approach of Galichon and Henry (2011) may be a preferred method. However, when the smallest CDC is tractable and $\mathcal{Z}$ is rich, the CDC approach is simpler. See Section 4.4 for a related discussion.

Next, consider imposing constraints on the outcome response function $d \mapsto Y_d^*$. For example, suppose $\mathcal{D} = \{d_1, \dots, d_{|\mathcal{D}|}\}$ is an ordered set and the outcome response function is monotonically increasing. This can be imposed by letting

$$\mathcal{S}^I_{Y^*} = \{(y_1, \dots . y_{|\mathcal{D}|}) \in \mathcal{Y}^{|\mathcal{D}|} : y_d \leqslant y_{d+1} \text{ for all } d = 1, \dots, |\mathcal{D}| - 1\}.$$

Additionally, concavity of $d \mapsto Y(d)$ can be imposed by letting:

$$\mathcal{S}^{IC}_{Y^*} = \mathcal{S}^I_{Y^*} \cap \{(y_1, \dots . y_{|\mathcal{D}|}) \in \mathcal{Y}^{|\mathcal{D}|} : y_{d+1} - y_d \geqslant y_{d+2} - y_{d+1} \text{ for all } d = 1, \dots, |\mathcal{D}| - 2\}.$$

These assumptions substantially restrict the outcome space, which results in a much smaller CDC. Panel (b) of Table 4 illustrates.

Finally, consider imposing more structure on the relationship between D and Z. Suppose that in addition to the potential outcomes Y^*, each unit in the population is characterized by a vector $D^* = (D^*_z)_{z\in\mathcal{Z}}$ of potential treatments. Suppose the observed treatment is $D = \sum_{z\in\mathcal{Z}} \mathbf{1}(Z = z)D^*_z$, and the instrumtent Z is jointly independent from (Y^*, D^*). Further, let $\mathcal{S} \subseteq \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|}$ summarize the restrictions on the outcome and treatment response functions. Then, given (Y, D, Z), the model produces a set of possible values 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 $B_d(Y) = (\mathcal{Y} \times \cdots \times \{Y\} \times \dots \mathcal{Y})$ with $\{Y\}$ in the d-th coordinate, and $B_z(D) = (\mathcal{D} \times \cdots \times \{D\} \times \dots \mathcal{D})$ with $\{D\}$ in the z-th coordinate. Note that 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 $G(y, d, z)$. After simplification, the Artstein's (in)equalities reduce to:

$$P(Y_d^* = y, D_z^* = d) = P(Y = y, D = d \,|\, Z = z),$$

for all $(y, d) \in \mathcal{S}$, $z \in \mathcal{Z}$. These equalities also follow directly from the assumed relationships between (Y, D), and (Y^*, D^*), and independence of the instrument:

$$P(Y = y, D = d \,|\, Z = z) = P(Y_d^* = y, D_z^* = d \,|\, Z = z) = P(Y_d^* = y, D_z^* = d),$$

as in Balke and Pearl (1997) or Bai, Huang, Moon, Shaikh, and Vytlacil (2024). ■

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

Example 5 (Directed Network Formation)**.** There are N firms forming 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. 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 $2^{2^{N(N-1)}}$, making 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 with complementarities Example 1, with $\delta_j > 0$. Then, the payoff from each local game depends only on the outcome of that local game, and the entire network Y is a PSNE if and only if the outcome Y^k of each of the local games Γ_k is a PSNE. 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 $\mathrm{Core}(G^1) \times \cdots \times \mathrm{Core}(G^N) = \mathrm{Core}(G) \cap \mathcal{S}$, 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 $\mathcal{S}$, $\Theta_0 \subseteq \Theta_0'$, because the latter only checks a subset of inequalities from the former.

To characterize Θ_0', Theorem 1 can be applied to each Γ_j 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. While the computational burden is lifted substantially, the resulting set of inequalities is still too large. To simplify the problem further, one may additionally impose the type-heterogeneity assumption. The details are left for further research. ■

4 Implementation and Relation to Other Methods

4.1 The Master Algorithm

The moment inequalities for characterizing Θ_0 as in Equation (8) can be obtained as follows.

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

1. ***Partition the parameter space***. *Partition the parameter space, $\Theta = \bigcup_{m=1}^{M} \Theta_m$, so that the support of $G(U, X; \theta)$ does not change with θ within each Θ_m. This step is application-specific and not always required, as discussed in Section 3.3. The partition can often be constructed analytically. For linear specifications, the partition can be obtained numerically using Algorithm 3 in Gu, Russell, and Stringham (2022).*

2. ***Partition the latent variable space***. *Fix $m \in \{1, \dots, M\}$ and any $\theta \in \Theta_m$. Let $\mathcal{Y} = \{y_1, \dots, y_s\}$ denote the outcome space and $S(x; \theta) = \{G_1, \dots, G_K\}$ denote set of all possible values of $G(U, x; \theta)$, conditional on $X = x$. Partition the latent variable space as $\mathcal{U}(x, \theta) = \{u_1, \dots u_K\}$, where $u_k = \{u \in \mathcal{U} : G(u, x; \theta) = G_k\}$, and define a measure $P_{(x,\theta)}$ on $\mathcal{U}_{(x,\theta)}$ by $P_{(x,\theta)}(u_k) = P(U \in u_k | X = x)$ for all $k = 1, \dots, K$. The probabilities $P_{(x,\theta)}$ can be computed by re-sampling or numerical integration.*

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

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

5. ***Compute the identified set.*** *Repeating Steps 2 – 4, compute the classes $\mathcal{C}_m^*(x)$ for all $x \in \mathcal{X}$ and $m = 1, \dots, M$ to obtain Θ_0.*

4.2 Computing the Smallest Core-Determining Class

Recall from Theorem 1 that the smallest CDC consists of the critical and implicit-equality sets. The latter can be easily 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 this end, suppose the graph $\mathbf{B}$ 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. Additionally, say that a set C is a *minimal critical superset* of A if there is no critical set $\tilde{C}$ such that $A \subset \tilde{C} \subset C$. Then, the main 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, the correspondence will satisfy $A \subseteq C$ for each $C \in F(A)$, and $F(\mathcal{Y}) = \varnothing$. For a collection of sets $\mathcal{C}$, define $F(\mathcal{C}) = \cup_{A \in \mathcal{C}} F(A)$. Then, the proposed algorithm simply iterates on F starting from the class $\mathcal{C} = \{G(u) : u \in \mathcal{U}\}$ until there are no more non-trivial 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 using the following algorithm.

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

Input: *A connected bipartite graph* $\mathbf{B}$ *and a self-connected set* A*.*
Output: *A 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* $P(C)$*.*

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

The algorithm 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* $k = 1, \dots, K$.

2. *For* $k = 1, \dots, K$*:*

   - *Initialize* $\mathcal{C}_k = \{G(u) : u \in \mathcal{U}_k\}$ *and* $R_k = \varnothing$.
   - *For each* $C \in \mathcal{C}_k$*: check if* 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* $\mathcal{C}_k$ *until convergence and collect all sets along the way into* R_k.

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

The formal proofs are given in Appendix A.4. Since at every iteration, except possibly the first one, 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. 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 where the CDC has cardinality less than a thousand, computation takes at most several seconds, even in settings where the total number of inequalities is prohibitively large and other existing algorithms are infeasible. Since the complexity is proportional to the size of the smallest CDC, further substantial improvements are not possible.

4.3 Additional Restrictions, Counterfactuals, and Inference

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 discussed in the examples in Section 3.3, additional restrictions on the model, such as independence of latent variables and excluded instruments, shape restrictions, support restrictions, and restrictions on the underlying selection mechanisms, can be easily accommodated. In some settings (recall Examples 1 and 5), these additional restrictions allow to simplify the analysis without losing sharpness.

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 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 4 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. As discussed in 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 in Torgovitsky (2019) and Gu, Russell, and Stringham (2022), is discussed in the next section.

Third, consider inference on θ_0. When the CDC does not change with θ (recall the discussion in Examples 1–4), standard inference procedures for moment inequalities apply; see, e.g., 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 uniformly valid test for $\theta \in \Theta_{0,m}(P)$, it is easy to verify that the test $\hat{\phi}_n(\theta) = \sum_{m=1}^{M} \hat{\phi}_{m,n}(\theta)\mathbf{1}(\theta \in \Theta_{0,m}(P))$ is uniformly valid for $H_0 : \theta \in \Theta_0(P)$.[6] Thus, to construct a uniformly valid confidence interval for θ_0, the practitioner may test $H_0 : \theta \in \Theta_{0,m}(P)$, for each $\theta \in \Theta_m$, for each $m = 1, \ldots, M$, and collect the points for which the test does not reject. Moreover, if the implicit equality sets differ across Θ_m, the above approach is expected to be more powerful in large samples than using all of Artstein's inequalities because it incorporates the information that certain Artstein's inequalities are binding. The existing procedures for subvector inference (see, e.g., Romano and Shaikh, 2008; Bugni, 2016; Kaido, Molinari, and Stoye, 2019) can also be modified to accommodate situations where the set of relevant moment inequalities depends on θ. Pursuing such modifications formally is beyond the scope of this paper.

Importantly, the CDC approach allows to find the inequalities that are redundant for identification. Since the CDC, by definition, does not depend on the observed distribution of the data, the inequalities redundant for identification are also redundant for point estimation of the identified set Θ_0 or bounds on the functionals $\phi(\theta_0)$. However, an important related question is whether these inequalities are also redundant for inference in finite samples. For example, consider the inequalities $P(Y \in A_j) \geqslant P(G(U;\theta) \in A_j)$, for $j \in \{0, 1, 2\}$, and

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

suppose A_0 is redundant given A_1, A_2 in the sense of Theorem 1. Intuitively, if $P(Y \in A_0)$ is estimated with higher precision than $P(Y \in A_1)$ and $P(Y \in A_2)$, it may be beneficial to include this inequality in the analysis.[7] This question arises more broadly in moment inequality models, and there are currently no available methods to decide when to use such "redundant" inequalities for inference. Developing a formal finite-sample criterion is an interesting direction for further research.

4.4 Comparison with Other Existing Approaches

Several alternative approaches are available for characterizing sharp identified sets in models with set-valued predictions. All of them require solving well-behaved optimization problems for each x and θ. This section describes each approach in more detail and compares it with CDC approach in terms of: (i) computational tractability; (ii) obtaining sharp bounds on functionals of the form $\phi(\theta_0)$ for some $\phi : \Theta \to \mathbb{R}$ known or identified from the data; and (iii) simplicity of 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 $P_{Y \mid X=x} = (P(Y = y \mid X = x))_{y \in \mathcal{Y}} \in [0,1]^{|\mathcal{Y}|}$ and $P_{(x;\theta)} = (P(U \in u \mid X = x))_{u \in \mathcal{U}} \in [0,1]^{|\mathcal{U}|}$.

4.4.1 Aumann Expectation via Support Function

Beresteanu, Molchanov, and Molinari (2011) represent the set $\mathcal{P}(x; \theta)$ as the Aumann expectation of a suitably defined random set. The Aumann expectation of a random set $Q(U, X; \theta) \subseteq \mathcal{Y}^* \subseteq \mathbb{R}^{d_{\mathcal{Y}^*}}$, denoted $\mathbf{E}[Q(U, X; \theta) \mid X]$, is defined as the set of conditional expectations of all of its measurable selections, so Y^* is a measurable selection of $Q(U, X; \theta)$ if and only if $\mathbb{E}[Y^* | X] \in \mathbf{E}[Q(U, X; \theta) | X]$ almost-surely. In non-atomic, or purely atomic, probability spaces, the Aumann expectation is always a convex set, so it can be characterized via the support function, $h_{\mathbf{E}[Q \mid X]}(t) = \sup_{a \in \mathbf{E}[Q \mid X]} a^T t$, defined on the unit ball $t \in B \subseteq \mathbb{R}^{|\mathcal{Y}^*|}$. While the Aumann expectation may be hard to compute in practice, one can show that $h_{\mathbf{E}[Q|X]}(t) = \mathbb{E}[h_Q(t)|X]$ holds almost-surely. If the expected support function is easy to compute, the sharp identified set can characterized by solving, for each θ and x, a concave maximization problem in $\mathbb{R}^{d_{\mathcal{Y}^*}}$:

$$\Theta_0 = \{\theta \in \Theta : \sup_{t \in B} (t^T \mathbb{E}[Y^* \mid X = x] - \mathbb{E}[h_{Q(U,x;\theta)}(t) \mid X = x]) \leqslant 0, \; x \in \mathcal{X} \text{ a.s.}\}. \tag{9}$$

[7] If the researcher suspects that this situation is empirically relevant, Theorem 1 can be applied "backwards" to generate the relevant redundant inequalities by adding up the critical ones (i.e., $A_0 = A_1 \cup A_2$). However, if such suspicion is based on the data, the inference procedures will have to be adjusted accordingly.

Beresteanu, Molchanov, and Molinari (2011) apply the above characterization to models with interval-valued outcomes and covariates and finite games with various solution concepts. In these settings, the Artstein's inequalities do not generally lead to tractable characterizations of the sharp identified sets.

This idea can also be used in discrete-outcome models. For each $y \in \mathcal{Y}$, define $y^*(y) = \mathbf{1}_y = (\mathbf{1}(\tilde{y} = y))_{\tilde{y} \in \mathcal{Y}}$, so that $y^* \in \mathcal{Y}^* = \{0,1\}^{|\mathcal{Y}|}$, and consider a random set $Q(U, X; \theta) = \{y^*(y) \in \mathcal{Y}^* : y \in G(U, X; \theta)\}$. Then, $Y \in G(U, X; \theta)$ if and only if $Y^* \in Q(U, X; \theta)$, by construction. With some algebra, the expected support function equals to

$$\mathbb{E}[h_{Q(U,X;\theta)}(t) \mid X = x] = \sum_{u \in \mathcal{U}} P_{(x;\theta)}(u) \max\{t^T \mathbf{1}_y : y \in G(u, x; \theta)\},$$

so the sharp identified set Θ_0 contains all $\theta \in \Theta$ such that:

$$\max_{t \in B} \left(t^T P_{Y|X=x} - \sum_{u \in \mathcal{U}} P_{(x;\theta)}(u) \max\{t^T \mathbf{1}_y : y \in G(u, x; \theta)\} \right) \leqslant 0, \quad x \in \mathcal{X} \text{ a.s..}$$

This maximization problem is concave and thus easy to solve even for large $|\mathcal{Y}|$ and $|\mathcal{U}|$.

For checking whether a given parameter value belongs to the sharp identified set, the above approach often remains computationally tractable when the smallest CDC is prohibitively large, providing 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 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 (9) 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).

4.4.2 Mixed Matching

Another way to characterize sharp identified sets is by verifying the existence of a mixed matching (Artstein, 1983; Galichon and Henry, 2011; Russell, 2021). Fix $x \in \mathcal{X}$. A mixed matching is the joint distribution $\pi(u, y, x; \theta)$ supported on $\mathrm{Gr}(G) = \{(u, y) \in \mathcal{U} \times \mathcal{Y} : u \in G(u)\}$, for which the $\mathcal{U}$-marginal matches $P_{(x;\theta)}$ and the $\mathcal{Y}$-marginal matches $P_{Y|X=x}$. Such joint distribution can be visualized by attaching a weight $\pi(u, y; x, \theta) \geqslant 0$ to every edge

$(u, y) \in E(\mathbf{B})$ and requiring:

$$\begin{aligned} \textstyle\sum_{u \in G^{-1}(y)} \pi(y, u; x, \theta) &= P_{Y|X=x}(y) \quad \text{for all } y \in \mathcal{Y}, \\ \textstyle\sum_{y \in G(u)} \pi(y, u; x, \theta) &= P_{(x;\theta)}(u) \quad \text{for all } u \in \mathcal{U}. \end{aligned} \tag{10}$$

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

$$\min_{\eta \in \mathbb{R}^{|\mathcal{Y}|+|\mathcal{U}|}} \left(b(x;\theta)^T \eta \,|\, A(x;\theta)^T \eta \geqslant 0 \right) \geqslant 0, \tag{11}$$

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 (10). So, the sharp identified set for θ can be characterized as:

$$\Theta_0 = \{\theta \in \Theta : (11) \text{ holds } x \in \mathcal{X}\text{-a.s.}\}. \tag{12}$$

Galichon and Henry (2011) propose an alternative optimal transport formulation of the problem. Consider a problem in which 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. Suppose the transportation cost is zero-one, where a pair (u, y) is assigned cost zero if $y \in G(u)$ and one otherwise. Then, the joint distribution $\pi(u, y; x, \theta)$ exists if and only if the optimal transport problem outlined above 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).[8]

The mixed matching approach often remains computationally tractable when the smallest CDC is not, providing 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 4 discussed in Section 3.3.[9] With the CDC approach, the conditional Artstein's inequalities can simply be intersected over Z. With the mixed matching approach, to ensure that the $\mathcal{U}$-marginal of π is independent of Z, additional $|\mathcal{Z}| - 1$ matching constraints are required, for each $u \in \mathcal{U}$. When $|\mathcal{Z}|$ is large or infinite, the task becomes infeasible. Further, if the counterfactual $\phi(\theta_0)$ of interest can be expressed directly in terms of π, the sharp bounds on it can be obtained by solving linear programs. For instance, in

[8] 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 minimzation 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 $|\mathcal{U}| \gg |\mathcal{Y}|^3$.

[9] To match the notation of this section and Example 4, let $U = Y^*$, $X = \varnothing$, and $Y = (Y, D)$.

the context of Example 4, Russell (2021) provides evidence that the linear programs scale favorably with $|\mathcal{Y}|$, for fixed $|\mathcal{D}|$ and $|\mathcal{Z}|$. More generally, like the support function approach, Equations (11)–(12) 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.

4.4.3 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 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 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 structure allows 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 ensuring 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. The "cells" in MRP 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 $P(Y \in A \,|\, X = x) \geqslant \sum_{k \in G^-(A)} \mu(u_k, x)$, 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.

4.4.4 Final Remarks

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

5 Extensions

This section extends the main results, Lemmas 1 and 2, and Theorem 1, to models in which the outcome variable has infinite support. This requires a more nuanced formal setup.

5.1 Formal 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. 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 $u \in \mathcal{U}$. Recalling the definitions of $G^-(A), G^{-1}(A)$ in (3), measurability requires $G^-(A), G^{-1}(A) \in \mathcal{F}$ for every $A \in \mathfrak{F}$. Denote $N(A) = G^{-1}(A) \backslash G^-(A)$. As before, let Core$(G)$ denote the set of distributions of all measurable selections of G. For 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 on $\mathcal{Y}$ absolutely continuous with respect to Q. Say that $A = B$, Q-a.s., if $Q((A \cap B^c) \cup (B \cap A^c)) = 0$, and similarly for P.

Throughout this section, we impose the following technical assumption.

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 to control measure-zero sets in the outcome space $\mathcal{Y}$. In econometric applications, it is often possible to find Q such that $\text{Core}(G) \subseteq \mathcal{M}_Q$. For example, if the outcome $Y \in \mathcal{Y}$ is discrete, Q can be taken 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 a Lebesgue measure. The researcher can also choose Q to explicitly restrict the set of selections of interest; see the discussion in Example 6 below. Part 2 of the assumption is a mild regularity condition requiring that the realizations $G(u)$ and the overlaps between sets $A \cap G(u)$, for P-almost-all $u \in \mathcal{U}$, can be "detected" by the measure Q. It allows to introduce a notion of connectedness of the random set similar to connectedness of the bipartite graph in Section 3.1. The core-determining classes are defined accordingly.

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

As in discrete settings, critical and implicit-equality sets play a crucial role in the analysis.

Definition 5.2 (Critical and Implict-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. Associate each set $A \subseteq \mathcal{Y}$ with an equivalence class $[A]$ defined via $A' \sim A$ if $A = A'$, Q-a.s., and $G^{-}(A) = G^{-}(A')$, P-a.s.. Say that A is critical if $\mathcal{M}_Q(\mathfrak{F}\backslash[A]) \neq \mathcal{M}_Q(\mathfrak{F})$. Say that A is an implicit-equality set if $\mu(A) = C_G(A)$ for any $\mu \in \mathcal{M}_Q \cap \text{Core}(G)$.*

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)**.** *Let Assumption 5.1 hold and $G : (\mathcal{U}, \mathcal{F}, P) \rightrightarrows (\mathcal{Y}, \mathcal{B}, Q)$ be a random closed set. Say that G is connected if for any $A \in \mathcal{B}$ with $Q(A) > 0$, it follows that $P(N(A)) > 0$.*

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. Finally, 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..*

The following example illustrates the concepts introduced above.

Example 6 (Dominated Selections and Connected Random Sets)**.** Let $\mathcal{U} = [0,1]^2$ and $\mathcal{Y} = [0,1]$ both endowed with Borel sigma-fields. Let $U = (U_1, U_2)$ be a pair of random variables supported within a set $S = \{(u_1, u_2) \in [0,1]^2 : u_1 \leqslant u_2\}$. Consider a random closed set $G : [0,1]^2 \rightrightarrows [0,1]$ defined by $G(U) = [U_1, U_2]$.

Depending on the distribution P, the random set G may have only continuous selections, or it may have a full menu of selections, including continuous, discrete, and mixed. For example, suppose P is any continuous distribution supported on S. Then, for example, $Y = (U_1 + U_2)/2$ is a continuous selection of G. Since in this case, G can be arbitrarily narrow with positive probability, it follows that $\text{Core}(G) \subseteq \mathcal{M}_\lambda$, where λ is the Lebesgue measure on $[0,1]$. Thus, Part 1 of Assumption 5.1 is satisfied with $Q = \lambda$. For part 2 of Assumption 5.1, notice that for P-almost all u, the set $G(u)$ has positive length, and so $Q(G(u)) > 0$. Further, consider, for example, $A = [a_1, a_2]$ for some $0 < a_1 < a_2 < 1$. Then $N(A) = \{(u_1, u_2) \in S : u_1 \in A, u_2 > a_2\} \cup \{(u_1, u_2) \in S : u_1 < a_1, u_2 \in A\}$. For P-almost all $u \in N(A)$, the segment $G(u) \cap A$ has positive length and thus $Q(G(U) \cap A) > 0$.

Alternatively, suppose $U_2 = U_1 + 1/K$, P-almost-surely, for some $K \in \mathbb{N}$. Then, for example, $Y = U_1$ is a continuous selection, and $Y = \sum_{k=0}^{K-1} \frac{k+1}{K} \mathbf{1}(U_1 \in \left[\frac{k}{K}, \frac{k+1}{K}\right))$ is a discrete selection of G. Thus, in this case, taking Q to be the Lebesgue measure on [0,1] will meaningfully restrict the set of selections of interest.

Next, suppose P is uniform on the union of sets $S_1 = S \cap \{(u_1, u_2) : u_2 \leqslant 1/2\}$ and $S_2 = S \cap \{(u_1, u_2) : u_1 \geqslant 1/2\}$. Then, $G^{-1}([0,1/2]) = S_1$ and $G^{-1}([1/2,1]) = S_2$. Then if Q is Lebesgue, the random set G is not connected in the sense of Definition 5.3. In this case, the restrictions $G_1 : S_1 \to [0,1/2]$ and $G_2 : S_2 \to [1/2,1]$ can be considered separately. ■

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 the outcome space $\mathcal{Y}$ 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 $A = \mathcal{Y}_n$ for some 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 that are the smallest by inclusion. Specifically, for each $l \geqslant 1$, the class $\bigcup_{n \geqslant 1} \mathcal{C}^*_n \cup \bigcup_{n \neq l} \mathcal{Y}_n$, where $\mathcal{C}^*_n$ is the set of all critical subsets of $\mathcal{Y}_n$ characterized in Lemma 3, is core-determining.*

Corollary 1.1 and the following discussion apply in continuous-outcome settings as well. 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. However, even when the sharp identified set for the full vector of parameters θ_0 is intractable, 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 examples below illustrate.

5.3 Examples

The first example discusses interval-valued outcomes; see Beresteanu, Molchanov, and Molinari (2012), Molinari (2020), and references therein.

Example 7 (Interval Data)**.** Let $Y^* \in \mathcal{Y}$ denote a continuously distributed outcome variable and $X \in \mathcal{X}$ denote covariates. Suppose the researcher does not observe Y^* directly, but has access to random variables $Y_L, Y_U \in \mathcal{Y}$ with $Y_L \leqslant Y_U$, almost surely, which are also continuously distributed and represent the lower and upper bounds on Y^* correspondingly. That is, $Y^* \in G(Y_L, Y_U) = [Y_L, Y_U]$. For simplicity, suppose X is discrete, and $\mathcal{Y} = [\underline{y}, \overline{y}]$ for some known $\underline{y}$ and $\overline{y}$. Additionally, assume that the joint distribution of (Y_L, Y_U), conditional on $X = x$, satisfies $P(\underline{\kappa}(x) \leqslant Y_U - Y_L \leqslant \overline{\kappa}(x) \,|\, X = x) = 1$, 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 $X = x$. Without loss of generality, Q can be taken to be the Lebesgue measure on $\mathcal{Y}$. Since for any $A \subseteq \mathcal{Y}$ with $Q(A) > 0$, $P(N(A)) > 0$, 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 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 $A = A_1 \cup A_2 = [a_1, b_1] \cup [a_2, b_2]$ 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, so such A is not self-connected. A similar argument applies to any other collection of disjoint intervals, meaning that all critical sets must be contiguous intervals. Next, consider an interval $A = [a, b]$ with $\underline{y} < a < b < \overline{y}$ and $b - a > \overline{\kappa}(x)$. 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., so such A is not complement-connected. Note that intervals of the form $[\underline{y}, b]$ and $[a, \overline{y}]$ are complement-connected. Therefore, the sharp identified set for θ_0 is fully characterized by the inequalities $P(Y^* \in A \,|\, X = x) \geqslant P([Y_L, Y_U] \subseteq A \,|\, X = x)$, 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 $\underline{\kappa}(x)$ or $\overline{\kappa}(x)$ are constant for all $x \in \mathcal{X}'$, for some $\mathcal{X}' \subseteq \mathcal{X}$, the corresponding inequalities can be intersected. Importantly, Theorem 2 implies that each of the above inequalities is also necessary to guarantee sharpness.

Now, 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 included in the "tube" of non-decreasing functions:

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

The upper and lower bounds are sharp in the sense of first-order stochastic dominance. However, not all CDF's inside the tube are included in the sharp identified set, because valid candidates must also satisfy the inequality

$$F_{Y^*|X=x}(b) - F_{Y^*|X=x}(a) \geqslant P(Y_L \geqslant a, Y_U \leqslant b | X = x) \tag{13}$$

for any a, b such that $\underline{\kappa}(x) \leqslant b - a \leqslant \overline{\kappa}(x)$. This rules out CDF's that stay constant or increase too little over any such interval. 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:

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

However, the upper bound may not be sharp due to (13) being violated at $a = q_{Y^*|X=x}(\tau_2)$,

$b = q_{Y^*|X=x}(\tau_1)$. Instead, one can 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 constructed similarly. ∎

The final example is an English auction model studied in Haile and Tamer (2003), Aradillas-López, Gandhi, and Quint (2013), Chesher and Rosen (2017), and Molinari (2020).

Example 8 (English 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 j-th smallest valuation and bid correspondingly. Let $F \in \mathcal{F}$ denote the joint distribution of ordered valutations $V = (V_{1:N}, \ldots, V_{N:N})$ supported on $S = \{v \in [0, \overline{v}]^N : v_1 \leqslant \cdots \leqslant v_N\}$. The set $\mathcal{F}$ summarizes the assumptions on the information structure. The researcher observes $B = (B_{1:N}, \ldots, B_{N:N})$ and wants to learn about some features of $\theta_0 = F$. For simplicity, suppose there is no reserve price and minimal bid increment. Further, 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 $V_{N-1:N} \leqslant B_{N:N}$. So, given the valuations V, the model produces a set of predictions for the bids:

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

Note that as long as the valuations are supported on S, the support of the random set $G(V; \theta)$ does not depend on θ. Thus, if the researcher has access to exogenous covariates Z independent from the valuations, the Artstein's inequalities can be intersected over the values of Z.

For simplicity, suppose $N = 2$, or only the top two bids are observed.[10] It is easy to verify that the random set $G(V; \theta)$ is connected, so in view of Lemma 4, there are no implicit-equality sets. In turn, the class of all critical sets is vast. It includes, for example, all lower sets $A_1 = \{(v_1, v_2) \in [0, \overline{v}]^2 : v_1 \leqslant \kappa(v_2)\}$, for some weakly decreasing function $\kappa : [0, \overline{v}] \to [0, \overline{v}]$, all sets of the form $A_2 = \{(v_1, v_2) : v_1 \leqslant a, v_2 \in [b, c]\}$, for some $a, b, c \in [0, \overline{v}]$ with $b \leqslant c$, 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 the joint distribution of ordered valuations, $\theta_0 = F$, is intractable.

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

[10] This is common in settings where it is hard to link the bids to the bidders, so one can only reliably use the top two bids.

that the expected profit and bidders surplus in English auctions depend only on the marginal distribution of the two largest valuations: $\phi(\theta_0) = (F_{N-1:N}, F_{N:N})$. The sharp identified set for $\phi(\theta_0)$ 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. This suffices to obtain closed-form sharp bounds on $F_{N-1:N}$ and $F_{N:N}$.

Specifically, an Artstein's inequality with the set $A = S \cap [0, v] \times [0, \overline{v}]$ implies an upper bound $F_{N-1:N}(v) \leqslant G_{N-1:N}(v)$, and the set $A = S \cap [0, \overline{v}] \times [v, \overline{v}]$ implies a lower bound $F_{N-1:N}(v) \geqslant G_{N:N}(v)$. Additionally, 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 imposed positive-dependence assumption on F yields the following bounds:

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

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, while the bounds on $F_{N-1:N}$ are sharp in the sense of first-order stochastic dominance, the corresponding "tube" of functions includes many CDF-s that do not belong to the sharp identified set. Specifically, the set $A = S \cap [a, \overline{v}] \times [0, b]$ for $b > a$ corresponds to the Artstein's inequality $F_{N-1:N}(b) - F_{N-1:N}(a) \geqslant P(B_{N-1:N} \geqslant a, B_{N:N} \leqslant b)$, which rules out the CDF-s 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 further research. ■

6 The Importance of Selecting Inequalities

This section provides 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 from Berry and Compiani (2020) and Example 2. In this setting, even with a few time periods, the total number of Artstein's inequalities is prohibitively large; see Table 2. To this end, the authors suggest using inequalities that should intuitively shrink the identified set. 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, I compare the resulting identified sets with the sharp identified set obtained using the smallest CDC.

I compare the identified sets in the model with $T = 2$ and $T = 5$ periods. The payoff parameters are set to $\pi = 0.5$, $\gamma = -1.5$, and $\rho = 0.75$. The sample size is 10,000 and the number of draws of the latent variables to approximate the containment functional in Artstein's inequalities is 300,000. Further details about the simulation routine are provided in Appendix B.

Figure 7 presents the results. The yellow sets correspond to the sharp identified sets in the model with $T = 2$, the blue sets combine the inequalities for $T = 2$ with the intuitive inequalities of Berry and Compiani (2020), and the red sets represent the sharp identified sets with $T = 5$. Evidently, the intuitive inequalities do not come close to utilizing all of the identifying information in the model with $T = 5$. In numerical terms, the blue (intuitive) identified set for (π, γ, ρ) is roughly 28% smaller than the yellow one, while the red (sharp) identified set is 93% smaller.

6.2 A Static Entry Model

In the second simulation exercise, we pose the following question: assuming the researcher uses some other inequality selection device, how likely are they to obtain an identified set approximately as informative as the sharp identified set?

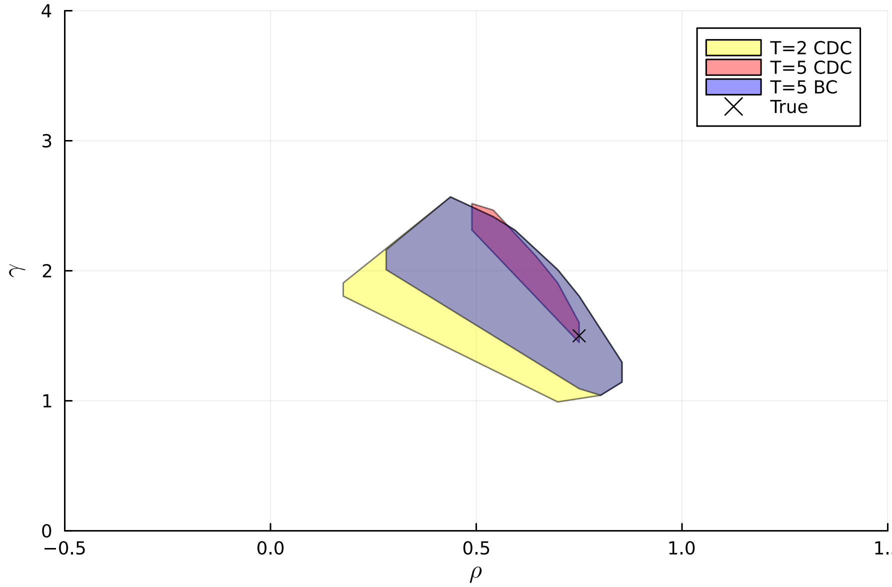
I revisit the market entry model from Example 1 with $N = 3$ players and strategic complementarities, $\delta_j > 0$ for all j. In this model, there are 254 non-trivial Artstein's inequalities, while the smallest CDC contains only 14 inequalities. Thus, a complete answer to the question posed above requires trying all combinations of 14 inequalities out of 254, which is computationally infeasible. To approximate such an experiment, we repeatedly sample 14 out of 254 inequalities at random and compute the corresponding identified sets using a fixed grid. Measuring the size of identified sets as the number of grid points satisfying the respective inequalities, we compute the ratio of the size of the sharp identified set to the size of the randomly generated one.

The parameters of the simulation are $\alpha_j = 1$ and $\delta_j = 1.75$ for all $j = 1, 2, 3$; the unobservables ε_j are distributed i.i.d. $N(0, 25)$. Within the regions of multiplicity, each of the equilibria is selected with equal probability. The sample size is 50,000.

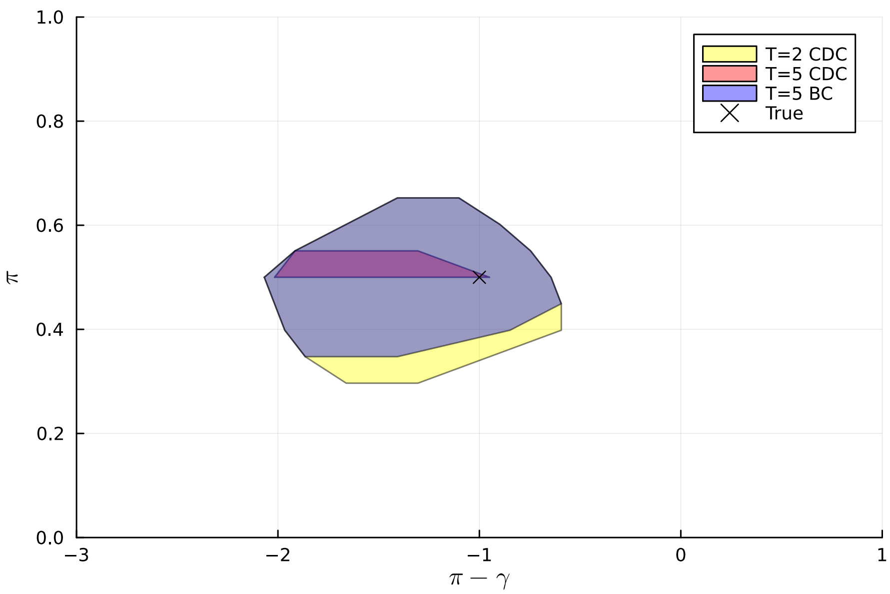
Figure 8 presents the results. Panels 8a–8c depict the sharp identified set and two examples of identified sets based on selecting other equally-sized collections of inequalities. The sharp identified set is relatively tight, while the identified sets based on other equally-sized collections of inequalities may be substantially larger. Panel 8d shows the distribution of the size of the sharp identified set relative to the simulated one across 200 simulations. For example, with probability 0.95, the sharp identified set is less than half of the non-sharp one, and with probability 0.8 it is less than a quarter. These results suggest that the CDC is a rather specific collection of inequalities in the sense that researchers using alternative and informal methods of inequality selection are likely to incur a substantial loss of identifying information.

7 Conclusion

A common practical problem in the analysis of partially-identified models with set-valued predictions is that the sharp identified sets are characterized by a very large number of moment inequalities. At the same time, many of those inequalities 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 information from the data and maintained assumptions. In this paper, we provided a simple characterization for the smallest possible core-determining class, illustrated its utility in several popular applications, and compared it with several alternative approaches. The results can be applied far beyond the class of examples considered in the paper. Determining what moment inequalities are more informative for inference in finite samples is a natural direction for further research.

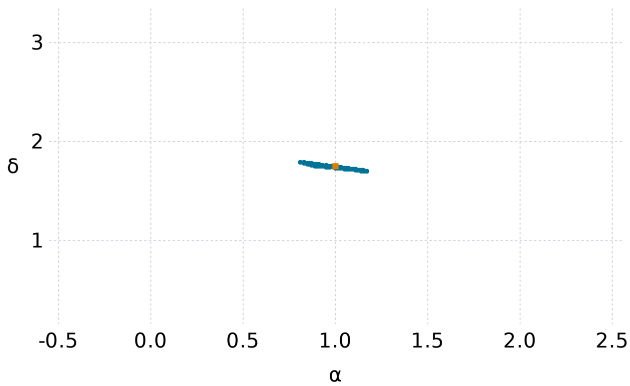


(a) Identified sets for the sunk cost γ and time dependence parameter ρ.

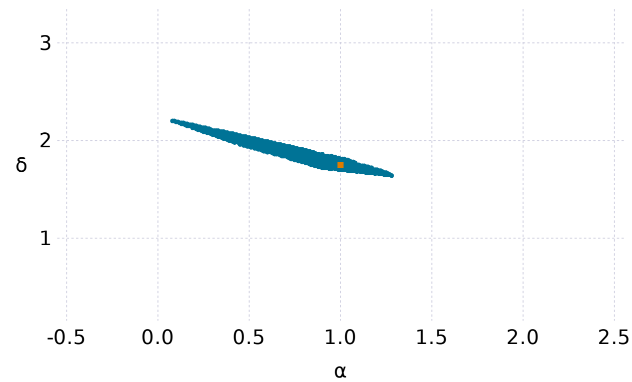


(b) Identified sets for the profit π and net profit $\pi - \gamma$.

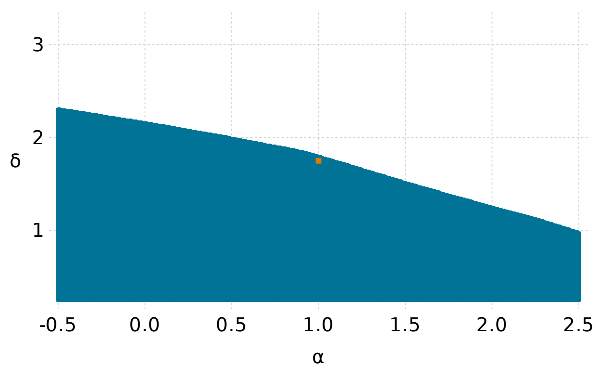
Figure 7: Identified sets in the dynamic entry model from Example 2.



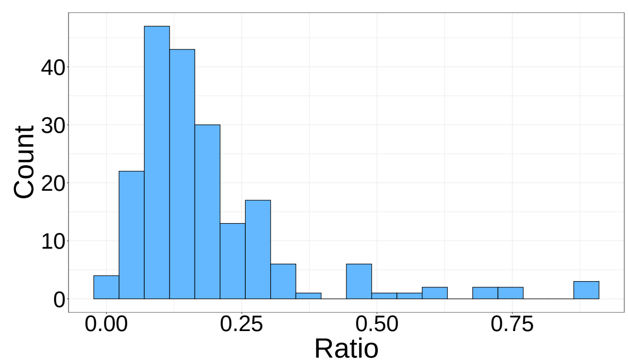
(a) Sharp identified set



(b) "Median" identified set



(c) "Unlucky" identified set



(d) Distribution of relative sizes of identified sets

Figure 8: 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.

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., 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. (2016): "Comparison of inferential methods in partially identified models in terms of error in coverage probability," *Econometric Theory*, 32(1), 187–242.

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*.

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.

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., 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. For the "If" direction, we show that for any set A that is both self- and complement-connected, there is a probability measure $\mu \in \text{Core}(G)$ satisfying $\mu(A) = C_G(A)$ and $\mu(\tilde{A}) > C_G(\tilde{A})$ for all $\tilde{A} \neq A$. Therefore, every such A must be 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 $\pi_0(u, \cdot)$ 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 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 $N(A) \neq \varnothing$, if follows that $\mu_0(A) > C_G(A)$, so A cannot be an implicit-equality set.

A.3 Theorem 1

First, suppose $\mathbf{B}$ is connected. To prove that the class $\mathcal{C}^*$ of all critical sets is core-determining, we show that the non-critical sets can be removed altogether without changing the core. Doing so would only lead to logical inconsistencies if there existed distinct non-critical sets A and B such that A must be used to argue that B is redundant and B must be used to argue that A is redundant. We will show that such sets cannot exist. 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 $G(u) \cap A^c \neq \varnothing$, for otherwise (i) is violated. But the existence of such u would contradict (ii).

Next, let $\mathcal{Y} = \bigcup_{m=1}^{M} \mathcal{Y}_m$ 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 $A = \cup_{m=1}^{M} A_m$ with $A_m \subseteq \mathcal{Y}_m$ satisfies $G^{-}(A) = \bigcup_{m=1}^{M} G^{-}(A_m)$, so it is redundant given $(A_m)_{m=1}^{M}$. Since $\sum_{m=1}^{M} \mu(\mathcal{Y}_m) = 1$ for any $\mu \in \text{Core}(G)$, any one (and only one) of the sets $\mathcal{M}$ can be omitted from the CDC. These facts, combined with the preceding argument, imply that the class of all critical subsets of all $\mathcal{Y}_m$ and all but one implicit-equality sets is the smallest CDC.

A.4 Algorithms 2 and 3

It suffices to show that Algorithm 2 finds all minimal critical supsersets 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 $C = A \cup B$, satisfying two properties: (i) C is self- and complement-connected; (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 $(\mathcal{Y}_l, \mathcal{U}_l, \mathcal{E}_l)$. Additionally, 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 following 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 satisfying $d\nu/dQ > 0$, 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 the Robbins' Theorem (Theorem 1.5.16 in Molchanov) 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 $\pi_0(u, \cdot)$ 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 goes through exactly as in the proof of 1 with all summations replaced by integrals and qualifiers P-a.s. and Q-a.s. added when referring to set operations in the $\mathcal{Y}$ and $\mathcal{U}$ spaces correspondingly.

A.6 Lemma 4 and Theorem 2

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

B Details for the Simulations in Section 6.

B.1 Dynamic Entry Game

The logic of the simulation follows that of Berry and Compiani (2020). Let T be the number of observed periods and $\bar{T} = 50 + T$ be the total number of periods used in the simulation. Let $N = 10{,}000$ be the sample size and $S = 300{,}000$ denote the number of latent variable draws.

The data is generated as follows. Draw N vectors ε of latent variables size $\bar{T}$ according to the $AR(1)$ process specified in Example 2. For each sample, draw $X_1 \sim \text{Bernoulli}(p = 0.5)$ and solve for the optimal policy for $\bar{T}$ periods. In period $\bar{T} - T$, for a randomly chosen half of the firms, replace $X_{\bar{T}-T}$ with $Z \sim \text{Bernoulli}(p = 0.5)$, and re-solve for the optimal policy from that point forward. This ensures that Z is a valid and relevant instrument for X_1. Keep the last T periods as the observed data.

There are three main parameters (π, γ, ρ) with the true values $(0.5, -1.5, 0.75)$, and an auxiliary parameter $\pi' = \pi - \gamma$. The grid has step size 0.05 and boundaries $\pi \in [-1.5, 1.5]$, $\pi' \in [-3, 0]$, and $\rho \in [0, 1]$.

C Additional Results

C.1 Size of the Smallest Core-Determining Class

Let $\mathcal{U} = \{u_1, \ldots, u_K\}$, $\mathcal{Y} = \{y_1, \ldots, y_S\}$, and $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ be a random set. Let $\mathcal{Y}$ be endowed with a total order $\succcurlyeq_{\mathcal{Y}}$ so that $y_1 \preccurlyeq_{\mathcal{Y}} \ldots \preccurlyeq_{\mathcal{Y}} y_S$. For a subset $A \subseteq \mathcal{Y}$, let $\underline{A}$ and $\overline{A}$ denote the least and the greatest elements of A with respect to $\succcurlyeq_{\mathcal{Y}}$. Say that the correspondence G is *contiguous with respect to* $\succcurlyeq_{\mathcal{Y}}$ if each $G(u_k)$ is of the form $\{y_m, y_{m+1}, \ldots, y_{m+l}\}$, for some m, l. Say that G is *marginally monotone* if there exists a total order $\succcurlyeq_{\mathcal{U}}$ on $\mathcal{U}$ such that $u_k \succcurlyeq_{\mathcal{U}} u_l$ implies $\underline{G}(u_k) \succcurlyeq_{\mathcal{Y}} \underline{G}(u_l)$ and $\overline{G}(u_k) \succcurlyeq_{\mathcal{Y}} \overline{G}(u_l)$. Say that G is *monotone* if there exist total orders $\succcurlyeq_{\mathcal{Y}}$ and $\succcurlyeq_{\mathcal{U}}$ such that G is contiguous and marginally monotone. With these definitions, a generic core-determining class can be obtained as follows.

Theorem C.1 (Discrete Random Sets in Totally Ordered Spaces)**.** *Let $\mathcal{U}$ and $\mathcal{Y}$ be finite sets, $\succcurlyeq_{\mathcal{Y}}$ a total order on $\mathcal{Y}$, and $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ a random set. Let $\mathcal{S}$ denote the class of all segments in $\mathcal{Y}$ that are contiguous with respect to $\succcurlyeq_{\mathcal{Y}}$. Define a total order $\succcurlyeq_{\mathcal{U}}$ on $\mathcal{U}$ by ordering first w.r.t. $\underline{G}(u) \succcurlyeq_{\mathcal{Y}} \underline{G}(u')$, then w.r.t. $\overline{G}(u) \succcurlyeq_{\mathcal{Y}} \overline{G}(u')$, and arbitrarily within equivalence classes. Let $\mathcal{U}_{NC} = \{u \in \mathcal{U} : G(u) \notin \mathcal{S}\}$ and $\mathcal{U}_{NM} = \{u \in \mathcal{U} : \exists u' \preccurlyeq_{\mathcal{U}} u : \underline{G}(u') \preccurlyeq_{\mathcal{Y}} G(u) \prec_{\mathcal{Y}} \overline{G}(u')\}$ collect the elements of $\mathcal{U}$ corresponding to non-contiguous (NC) or non-monotone (NM) values of $G(u)$. Let $\mathcal{A}_{\varnothing} = \{S \in \mathcal{S} : G^{-}(S) = \varnothing\}$, and $\mathcal{A}_{NM} = \{G(u) : u \in \mathcal{U}_{NM}\}$. Let $\mathcal{C}$ denote the class of sets S such that:*

1. *$S = \bigcup_{l=1}^{L} S_l$, where $S_l \in \mathcal{S}$, $\underline{S}_{l+1} \succ_{\mathcal{Y}} \overline{S}_l$, and $A_l = (\overline{S}_l, \underline{S}_{l+1}) \neq \varnothing$.*
2. *Each interior S_l and A_l is in $\cup_{\mathcal{A}_{NM} \cup \mathcal{A}_{\varnothing}}$.*
3. *If $L \geqslant 2$, $G^{-}(S) \cap \mathcal{U}_{NC} \neq \varnothing$.*

Then, $\mathcal{C}$ is core-determining. The smallest core-determining class can be computed by applying Theorem 1 to $\mathcal{C}$.

Despite a lengthy statement, the idea of Theorem C.1 is simple: when $\mathcal{Y}$ is ordered, any subset $S \subseteq \mathcal{Y}$ can be expressed as a union of disjoint contiguous segments, and for such S to be non-redundant, these contiguous segments and the "gaps" between them must satisfy certain conditions. The formal proof is provided at the end of this section. Theorem C.1 implies that the smallest CDC will have a relatively small cardinality when the sets $\mathcal{U}_{NC}$ and $\mathcal{U}_{NM}$ of elements that violate either contiguity or marginal monotonicity are empty or contain only a few elements. An immediate corollary is the following.

Corollary C.1.1. *Let $\mathcal{U}$ and $\mathcal{Y}$ be finite sets and $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ a random set. If there exists a total order $\succcurlyeq_{\mathcal{Y}}$ such that G is contiguous, the size of the smallest CDC is of order $|\mathcal{Y}|^2$, at most. If, additionally, G is monotone, the smallest CDC has a cardinality of order $|\mathcal{Y}|$.*

The second part of the statement is essentially Theorem 4 in Galichon and Henry (2011). While contiguity and monotonicity are almost never satisfied in practice, one may expect that if G is contiguous (resp. monotone) "most of the time," the size of the smallest core-determining class should not be much larger than $|\mathcal{Y}|^2$ (resp. $|\mathcal{Y}|$). The notions of monotonicity and contiguity depend on the chosen order $\succcurlyeq_{\mathcal{Y}}$. Examples suggest that the most suitable order is not unique and hard to guess or describe analytically. However, it can be easily found numerically, using the following algorithm.

Algorithm 4 (Graph Rearrangements)**.**

Input: *a bipartite graph $\mathbf{B}_G$ with vertices $(\mathcal{Y},\mathcal{U})$, either connected or decomposed into connected components. The algorithm below applies to each of the connected components.*

1. *Calculate all pairwise distances between the vertices in $\mathcal{Y}$ and find all pairs $(y_m, y'_m)_{m=1}^M$ that are the most distant in $\mathbf{B}_G$.*
2. *For each $m = 1, \ldots, M$, apply the greedy algorithm:*
   (1) *Let $y_m^1 = y_m$ and for each $l \geqslant 1$, pick an element y_m^{l+1} with the smallest betweenness centrality from the set*
   $$\operatorname*{argmin}_{y \notin \mathcal{Y}_m^{1:l}} d(y, y_m^l),$$
   where $\mathcal{Y}_m^{1:l} = \{y_m^1, \ldots, y_m^l\}$. If there are ties, consider all possible continuations.
   (2) *Among all orders obtained in the previous step, pick any order with the smallest weight $\sum_{l=1}^{S-1} d(y_m^l, y_{m+1}^l)$. Denote the selected order $\succcurlyeq_{\mathcal{Y}}^m$.*
   (3) *Reorder the elements of $\mathcal{U}$ so that $u \succcurlyeq_{\mathcal{U}}^m u'$ if $\min G(u) \succcurlyeq_{\mathcal{Y}}^m \min G(u')$ and $\max G(u) \succcurlyeq_{\mathcal{Y}}^m \max G(u')$ and let $\mathbf{B}_G^m$ denote the resulting bipartite graph.*
3. *Among all $\{\mathbf{B}_G^m : m = 1, \ldots, M\}$, select the "most monotone one" in the sence defined below.*

Let $\mathbf{B}$ denote a bipartite graph with vertices in $\mathcal{U} = \{u_1, \ldots, u_K\}$ and $\mathcal{Y} = \{y_1, \ldots, y_S\}$ and edges $\mathcal{E}(\mathbf{B})$. The *edge list* of $\mathbf{B}$ is a set $\mathbf{E} = \{(k, s) : (u_k, y_s) \in \mathcal{E}(\mathbf{B})\}$, so that $\mathbf{E} \subseteq \{1, \ldots, K\} \times \{1, \ldots, S\}$. Denote the slices of $\mathbf{E}$ by $\mathbf{E}_k = \{s : (k, s) \in \mathbf{E}\}$ and $\mathbf{E}_s = \{k : (k, s) \in \mathbf{E}\}$. Let $\mathcal{I}(\mathbf{B})$ denote the set of all bipartite graphs $\mathbf{B}'$ isomorphic to $\mathbf{B}$, that is, the graphs obtained from $\mathbf{B}$ by permuting the vertices $\mathcal{Y}$ and $\mathcal{U}$ while keeping adjacency structure constant. Let $\mathbf{B}'$ denote a generic element of $\mathcal{I}(\mathbf{B})$ and $\mathbf{E}'$ the corresponding edge list. Define a partial order $\succcurlyeq_C$ on $\mathcal{I}(\mathbf{B})$ as:

$$\mathbf{B}' \succcurlyeq_C \mathbf{B} \iff \begin{cases} \min_s \mathbf{E}'_k, \geqslant \min_s \mathbf{E}_k & \text{for all } k = 1, \ldots, K \\ \max_s \mathbf{E}'_k \leqslant \max_s \mathbf{E}_k & \\ \min_k \mathbf{E}'_s \geqslant \min_k \mathbf{E}_s & \text{for all } s = 1, \ldots, S \\ \max_k \mathbf{E}'_s \leqslant \max_k \mathbf{E}_s & \end{cases}$$

where the inequalities are interpreted component-wise.

C.1.1 Proof of Theorem C.1

Below, we refer to segments contiguous with respect to $\succcurlyeq_{\mathcal{Y}}$ simply as segments. Say that a segment $S \subseteq \mathcal{Y}$ is interior if it contains neither $\underline{\mathcal{Y}}$ nor $\overline{\mathcal{Y}}$. To show that $\mathcal{C}$ is a core-determining class, we will show that $\mathcal{C}^* \subseteq \mathcal{C}$, where $\mathcal{C}^*$ is the smallest core-determining class characterized in Corollary 1. First, consider an arbitrary interior segment $S \in \mathcal{C}^*$. It must be, in particular, that the subgraph $(S^c, G^{-1}(S^c))$ is connected. Then, there is a $u' \in \mathcal{U}$ such that $\underline{G}(u') \prec_{\mathcal{Y}} S \prec_{\mathcal{Y}} \overline{G}(u')$. Therefore, if $G(u) \subseteq S$, it must be that $u \in \mathcal{U}_{NM}$, meaning that $G^{-}(S) \subseteq \mathcal{U}_{NM}$. Additionally, $S \in \mathcal{C}^*$ implies that $S \in \mathsf{U}_G$. Combining the above observations yields $S \in \mathsf{U}_{\mathcal{A}_{NM}}$. Next, note that an arbitrary subset $S \subseteq \mathcal{Y}$ can be written as a union of disjoint segments. That is, $S = \bigcup_{l=1}^{L} S_l$, where $S_l \in \mathcal{S}$, $\underline{S}_{l+1} \succ_{\mathcal{Y}} \overline{S}_l$ and $A_l = (\overline{S}_l, \underline{S}_{l+1}) \neq \varnothing$. If $S \in \mathcal{C}^*$, the subgraphs induced by $(S, G^{-}(S))$ and $(S^c, G^{-}(S^c))$ must be connected. For any interior segment S_l, the argument from the preceding paragraph implies $G^{-}(S_l) \subseteq \mathcal{U}_{NM}$, possibly with $G^{-}(S_l) = \varnothing$. The same applies to each A_l, with the addition that connectedness of $(S, G^{-}(S))$ requires $G^{-}(S) \cap \mathcal{U}_{NS} \neq \varnothing$. Therefore, the class $\mathcal{C}$ defined in the statement of the theorem is core-determining.