

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

Kirill Ponomarev

Department of Economics, University of Chicago

kponomarev@uchicago.edu

This version: April 15, 2024

Abstract

One of the main challenges in partially-identified models is obtaining a tractable characterization of the sharp identified set, which exhausts all information contained in the data and modeling assumptions. In a large class of models, the sharp identified sets can be described using the so-called Artstein's inequalities, which verify that the observed distribution of outcomes, given covariates, can be generated by the model for some parameter values. While theoretically convenient, this approach is often impractical because the total number of inequalities is too large. However, many of the inequalities may be redundant in the sense that excluding them from the analysis does not lose identifying information. In this paper, I characterize the smallest possible collection of inequalities that suffices to describe the sharp identified set and provide an efficient algorithm for obtaining these inequalities in practice. I apply the results to the models of static and dynamic market entry, discrete choice, selectively-observed data, and English auctions, and conduct a simulation study to demonstrate that the proposed method substantially improves upon ad hoc inequality selection.

1 Introduction

Many partially-identified 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 potential values for the outcome $Y \in \mathcal{Y}$. The researcher does not observe $G(U, X; \theta)$ but postulates that $Y \in G(U, X; \theta_0)$ 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 left completely unspecified or somehow restricted. 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., [De Paula, Richards-Shubik, and Tamer, 2018](#); [Miyauchi, 2016](#); [Sheng, 2020](#); [Gualdani, 2021](#)), and English auctions (e.g., [Haile and Tamer, 2003](#); [Aradillas-López, Gandhi, and Quint, 2013](#)).

In a closely-related class of models, the roles of Y and U are switched: the researcher can compute a set $G(Y, X; \theta)$ for any θ , and assumes that $U \in G(Y, X; \theta_0)$ for some “true” parameter value θ_0 . Examples of such settings include models with missing or interval-valued 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](#)), semiparametric discrete choice models (e.g., [Chesher, Rosen, and Smolinski, 2013](#); [Chesher and Rosen, 2017](#); [Torgovitsky, 2019](#); [Tebaldi, Torgovitsky, and Yang, 2019](#)), and discrete choice models with heterogeneous or counterfactual choice sets ([Manski, 2007](#); [Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021](#)), among others.

The sharp identified sets for parameters θ in such models can be characterized as follows.¹ 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 $\{Y \in A\}$, so at $\theta = \theta_0$, the inequalities

$$P(Y \in A | X = x) \geq P(G(U, X; \theta) \subseteq A | X = x; \theta) \quad (1)$$

must hold for all measurable subsets $A \subseteq \mathcal{Y}$, all $x \in \mathcal{X}$. Thus, a natural identified set for θ is:

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

¹For brevity of exposition, I focus on models of the first type, but all arguments and results extend to models of the second type.

The results of Artstein (1983) imply that the above inequalities exhaust all information contained in the data and assumption $Y \in G(U, X; \theta_0)$, so Θ_0 is sharp. The inequalities in (1) are known as Artstein's inequalities.

The above characterization is theoretically appealing, but the total number of inequalities characterizing Θ_0 is often very large so verifying them all is infeasible in practice. In such cases, it is customary to select a smaller collection of inequalities by hand and proceed with an outer set for Θ_0 . While this approach may still deliver informative results, it has several important drawbacks. First, it is usually not clear *ex ante* which of the inequalities are relatively more informative, and ad hoc inequality selection may lead to substantial loss of information.² Second, relying on outer sets may lead to misleading conclusions if the model is misspecified. Specifically, Kédagni, Li, and Mourifié (2020) show that if the model is misspecified, in the sense that the sharp identified set is empty, one can always obtain two non-overlapping outer sets by selecting different subsets of moment inequalities.

At the same time, numerous examples suggest that many (even most) inequalities in (2) may be redundant in the sense that excluding them from the analysis does not change the resulting identified set. By identifying and removing such inequalities, it is often possible to make the analysis tractable while avoiding the loss of information and potentially misleading relaxations. In this paper, I propose a transparent and computationally efficient way to do so.

To address inequality selection, I focus on core-determining classes; see Galichon and Henry (2011); Chesher and Rosen (2017); Luo and Wang (2018); Molchanov and Molinari (2018). A class of \mathcal{C} of subsets of \mathcal{Y} is called *core-determining* if verifying (1) for all $A \in \mathcal{C}$ suffices to conclude that it holds for all $A \subseteq \mathcal{Y}$. Such \mathcal{C} may be considerably smaller than the class of all measurable subsets of \mathcal{Y} , resulting in a more tractable characterization of the sharp identified set. To obtain a core-determining class, it is necessary to identify redundant inequalities. In this paper, I propose a new analytical criterion to determine if a particular inequality is redundant, and use it to derive the smallest possible core-determining class. I show that this class is often considerably smaller than the classes existing in the literature, and, in discrete settings, provide an efficient way to compute it by checking connectivity of suitable subgraphs of a bipartite graph, representing the model's correspondence.

This paper contributes to the large and growing literature on econometrics with

²This point is illustrated in Section 5.

partial identification; see [Molinari \(2020\)](#); [Chesher and Rosen \(2020\)](#); [Pakes, Porter, Ho, and Ishii \(2015\)](#), and [Kline, Pakes, and Tamer \(2021\)](#) for detailed reviews. To best place this paper in the literature, the following high-level description of the identification problem will be useful (see [Matzkin, 2007](#)). Let $P(x; \theta, \eta)$ denote the model-implied conditional distribution of endogenous variables Y , given covariates $X = x$, and a parameter vector $(\theta, \eta) \in \Theta \times \mathcal{N}$. Suppose the vector θ represents the parameters of interest, and η represents the nuisance parameters. Setting $\mathcal{P}(x; \theta) = \{P(x; \theta, \eta) : \eta \in \mathcal{N}\}$, the sharp identified set for θ_0 is given by $\Theta_0 = \{\theta \in \Theta : P_{Y|X=x} \in \mathcal{P}(x; \theta), \forall x \in \mathcal{X}\}$. In this notation, all of the existing approaches to identification are based on obtaining a tractable characterization of the set $\mathcal{P}(x; \theta)$. In the models with set-valued predictions, the nuisance vector η includes a selection mechanism that chooses an outcome Y from the set of outcomes $G(U, X; \theta)$, predicted by the model. In models with point-valued predictions, η typically includes the conditional distribution of latent variables, given observables.

The most closely related papers, which also use Arstein's inequalities for identification, are [Galichon and Henry \(2011\)](#), [Chesher and Rosen \(2017\)](#) and [Luo and Wang \(2018\)](#). [Galichon and Henry \(2011\)](#) characterize the sets $\mathcal{P}(x; \theta)$ and sharp identified sets in discrete games using submodular optimization and optimal transport. To gain tractability, they introduce the notion of core-determining classes and show that if the model's correspondence is suitably monotone, there is a core-determining class whose size scales linearly with the size of the outcome space. While this result is powerful, it only applies in very restricted settings. More generally, the size of smallest core-determining class may scale exponentially with the size of the outcome space and is much harder to characterize. In this paper, I relate the smallest core-determining class to the structure of the underlying model's correspondence without any restrictions and propose an efficient algorithm to compute it in practice. [Chesher and Rosen \(2017\)](#) derive analytical sufficient conditions for identifying redundant inequalities of the form (1). By adding an extra step to their construction, I provide a necessary and sufficient condition for redundancy and use it to characterize the smallest possible core-determining class. In discrete settings, the proposed criterion amounts to checking if suitable subgraphs of the bipartite graph representing the model's correspondence are connected. A similar result in discrete settings appears in [Luo and Wang \(2018\)](#), but the approach proposed here applies more generally and leads to more efficient computation.

Two other closely related papers are [Beresteanu, Molchanov, and Molinari \(2011\)](#) and [Mbakop \(2023\)](#). [Beresteanu, Molchanov, and Molinari \(2011\)](#) study discrete games under various solution concepts and characterize the set $\mathcal{P}(x; \theta)$ as the Aumann expectation of a suitably defined random set. They proceed to characterize it via support functions, 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. In principle, both of the above approaches are applicable to some of the models studied in this paper, but as discussed in Section 4, they do not lead to computational improvements.

Other related work includes [Tebaldi, Torgovitsky, and Yang \(2019\)](#) and [Gu, Russell, and Stringham \(2022\)](#). The former paper focuses on discrete choice models with endogeneity, and the latter covers general discrete-outcome models. The authors 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 linear constraints on the cell probabilities, the sharp bounds on the counterfactual can be obtained using linear programming. In this paper, I show that suitable applications of Artstein's inequalities may help reduce the complexity of the resulting linear programs, making the two approaches complementary.

The rest of the paper is organized as follows. Section 2 presents motivating examples and provides 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 illustrates the utility of selecting inequalities in a simulation. Section 6 concludes.

2 Models with Set-Valued Predictions

2.1 Motivating Examples

To outline the scope of the paper, I consider four stylized examples. For a more detailed discussion, see [Molinari \(2020\)](#), [Chesher and Rosen \(2020\)](#), and [Kline, Pakes,](#)

and Tamer (2021). The first example is a static entry game studied in Tamer (2003), Ciliberto and Tamer (2009), and Beresteanu, Molchanov, and Molinari (2011), among others.

Example 1 (Static Entry Game). There are N firms, and each firm decides whether to stay out of enter the market, $Y_j \in \{0, 1\}$. The payoffs are given by:

$$\pi_j(Y, \varepsilon_j; \theta) = Y_j(\alpha_j + \delta_j(N(Y) - 1) + \varepsilon_j),$$

where $Y = (Y_1, \dots, Y_N)$ is the outcome vector, $N(Y)$ is the total number of entrants, $U = (\varepsilon_1, \dots, \varepsilon_N)$ are idiosyncratic payoff shifters, and $(\alpha_j, \delta_j)_{j=1}^N$ are payoff parameters. The firms have complete information and play a pure-strategy Nash Equilibrium. Market- and firm-level covariates can be accommodated by letting $(\alpha_j, \delta_j) = (\alpha_j(X), \delta_j(X))$ and specifying the relationship between X and U , but are omitted here for simplicity. The joint distribution of latent variables U , denoted by F , may be known exactly, or up to a parameter value, or left completely unspecified. Note that the model is incomplete: multiple equilibria are possible, but the equilibrium selection mechanism is not specified.

In this example, $\mathcal{Y} = \{0, 1\}^N$, $\mathcal{U} = \mathbb{R}^N$, and $\theta = (\alpha, \delta, F) \in \mathbb{R}^{2N} \times \Delta(\mathbb{R}^N)$. The set-valued prediction corresponds to the set of pure-strategy Nash Equilibria:

$$G(U; \theta) = \{Y \in \mathcal{Y} : Y_j = \mathbf{1}(\alpha_j + \delta_j(N(Y) - 1) + \varepsilon_j \geq 0), \text{ for all } j = 1, \dots, N\}.$$

Figure 1 illustrates possible realizations of $G(U; \theta)$ when $N = 2$ and $\delta_j < 0$ for $j = 1, 2$. With N firms, the size of the outcome space is 2^N , so there are $2^{2^N} - 2$ nontrivial moment inequalities of the form in (1). ■

The second example is a discrete choice model with endogeneity, studied in Chesher, Rosen, and Smolinski (2013) and Tebaldi, Torgovitsky, and Yang (2019), among others.

Example 2 (Discrete Choice with Endogeneity). Individuals choose one of $J + 1$ alternatives, $Y \in \{y_0, y_1, \dots, y_J\}$. Choosing y_j yields utility $v_j(X) + \varepsilon_j$, where $X \in \mathcal{X}$ includes prices, individual-level, and market-level covariates, and ε_j are latent utility shifters. Normalize $v_0 = 0$ and $\varepsilon_0 = 0$. Individuals know the utility of each alternative and choose the one that yields maximum utility, $Y = \operatorname{argmax}_j \{v_j(X) + \varepsilon_j\}$. Some

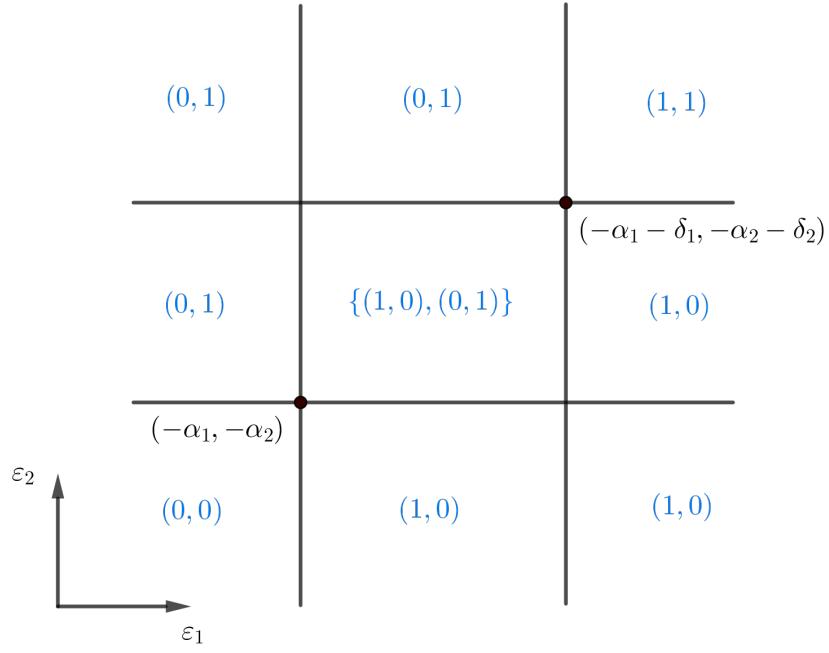


Figure 1: Set-valued prediction in a static entry game from Example 1 with $N = 2$ and $\delta_j < 0$.

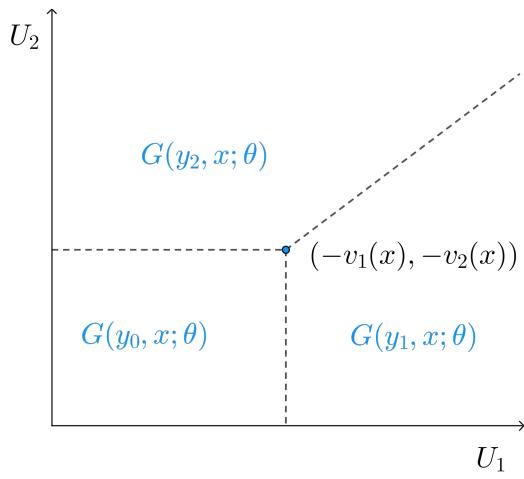


Figure 2: Set-valued prediction in a discrete choice model from Example 2 with $J = 2$.

components of X (e.g., prices) may be correlated with the latent payoff shifters $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_J)$ but the nature of this dependence is left unspecified. In this sense, the model is incomplete. The econometrician observes Y and X , and has access to instrumental variables $Z \in \mathcal{Z}$ such that ε is independent of Z .

This is an example of a model that delivers a set-valued prediction for latent variables, given the observables. Note that given $X = x$, the event $Y = y_j$ happens if and only if $v_j(x) + \varepsilon_j \geq v_k(x) + \varepsilon_k$ for all $k \neq j$. Denoting $U_j \equiv \varepsilon_j - \varepsilon_0$ for all $i \neq j$ and $U = (U_1, \dots, U_J) \in \mathbb{R}^J$, define:

$$G(y_j, x; \theta) = \begin{cases} \{u : u_j - u_k \geq v_k(x) - v_j(x), \text{ for all } k \neq j\} & j \geq 1 \\ \{u : u_k < -v_k(x), \text{ for all } k \geq 1\} & j = 0. \end{cases}$$

where $\theta = (v_1, \dots, v_M, \{F_{U|X=x}\}_{x \in \mathcal{X}})$. Then, the model can be summarized by:

$$\begin{aligned} U &\in G(Y, X; \theta), \text{ a.s.} \\ F_{U|Z=z}(u) &= F_U(u), \text{ for all } u \in \mathcal{U}, w \in \mathcal{W}, z \in \mathcal{Z} \end{aligned}$$

Figure 2 illustrates possible values $G(Y, x; \theta)$ for a given value of $X = x$ when $J = 2$. The sharp identified set can generally be described by a finite number of conditional moment inequalities of the form (1), for each value of the conditioning variables. Depending on the setup, one may choose to condition on both X and Z , as [Tebaldi, Torgovitsky, and Yang \(2019\)](#), or only on Z , as [Chesher, Rosen, and Smolinski \(2013\)](#). As explained in the sequel, these approaches are equivalent. ■

The third example discusses identification of features of the joint distribution of a partially-observed outcome and fully-observed covariates. This setting includes models with missing or censored outcome data as well as potential outcome models; For the application of partial identification methods in such models, see [Manski \(2003\)](#) and [Molinari \(2020\)](#), and references therein.

Example 3 (Partially Observed Outcomes). Let $Y^* \in \mathcal{Y}$ denote a partially-observed outcome variable, $X \in \mathcal{X}$ denote explanatory variables. Additionally, the econometrician observes a vector $W \in \mathcal{W}$, which may overlap with X , and can construct a set $G(W) \subseteq \mathcal{Y}$ such that $Y^* \in G(W)$ almost surely.

In the context of censored data, $W = (Y_L, Y_U)$ represent the observable bounds on the outcome of interest, and $G(W) = [Y_L, Y_U]$ so that $Y^* \in [Y_L, Y_U]$ almost

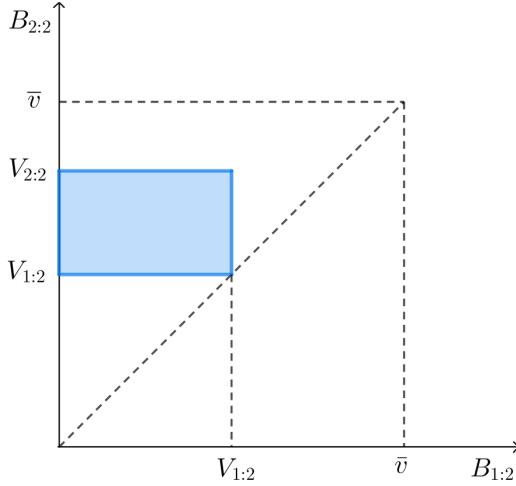


Figure 3: Set-valued prediction in an English auction with $N = 2$.

surely. Additional assumptions on the distribution of outcomes within the brackets, can be accommodated. In the context of missing data, $W = (D, Y)$, where $D \in \{0, 1\}$ indicates that the outcome $Y^* = Y$ is observed, and it is only known that $Y^* \in \mathcal{Y}$ otherwise. Denoting $G(W) = D\{Y\} + (1 - D)\mathcal{Y}$, by construction, $Y^* \in G(W)$ almost surely. Additional assumptions, specifying the relation between the distributions of $Y^*|D = 1$ and $Y^*|D = 0$, can be accommodated. In the standard potential outcomes model, $W = (D, Y)$, where $D \in \{0, 1\}$ denotes the treatment indicator, $Y^* = (Y_0^*, Y_1^*) \in \mathcal{Y}^2$ is a vector of potential outcomes, and $Y = DY_1^* + (1 - D)Y_0^* \in \mathcal{Y}$ is the observed outcome. Denoting $G(W) = D(\mathcal{Y} \times \{Y\}) + (1 - D)(\{\mathcal{Y}\} \times \mathcal{Y})$, by construction, $Y \in G(W)$ almost surely. Additional assumptions, such as monotonicity or exclusion restrictions, can be accommodated.

Parameters of interest include features of the distribution or quantile function of Y^* conditional on X , linear or quantile regression coefficients, and various treatment effects. Sharp identified sets can generally be characterized by an infinite number of moment inequalities of the form in (1). ■

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

Example 4 (English Auctions). Consider a symmetric ascending auction with N bidders. For simplicity, suppose that there is no reserve price and minimal bid increment. Let $V_j \in [0, \bar{v}]$ and $B_j \in [0, \bar{v}]$ 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 denote the joint distribution of ordered valuations $V = (V_{1:N}, \dots, V_{N:N})$ supported on $S = \{v \in [0, \bar{v}]^N : v_1 \leq \dots \leq v_N\}$. The researcher observes $B = (B_{1:N}, \dots, B_{N:N})$ and wants to learn about some features of $\theta_0 = F$. It is assumed 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 $B_{j:N} \leq V_{j:N}$ for all j , and (ii) implies $V_{N-1:N} \leq B_{N:N}$, so that the set-valued prediction is given by

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

Figure 3 presents an example realization of $G(V; \theta)$ with $N = 2$. In this setting, the sharp identified set for θ_0 is characterized by an infinite number of moment inequalities of the form in (1). ■

2.2 Sharp Identification via Artstein's Inequalities

In the above examples, the set of outcomes (or latent variables) predicted by the model depends on a realization of certain random variables, so it is a random set. Naturally, identified sets in such settings can be conveniently described using the language of random sets. I 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, $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ a standard Borel space (such as \mathbb{R}^d), and $\mathcal{P}(\mathcal{Y})$ denote the set of all probability measures on $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$. A *random closed set* is a measurable correspondence $G : \mathcal{U} \rightrightarrows \mathcal{Y}$ such that $G(u)$ is closed for all $u \in \mathcal{U}$.³ For each $A \subseteq \mathcal{Y}$, denote the upper and lower inverse of G by:

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

and note that $G^-(A) \subseteq G^{-1}(A)$. Let $S(G) = \{G(u) : u \in \text{supp}(P)\}$ denote the support of G and \mathbf{U}_G denote the set of all measurable unions of elements of $S(G)$.

³Measurability requires $G^-(A) \in \mathcal{F}$ for every closed set $A \subseteq \mathcal{Y}$, where $G^-(A)$ is defined in (3).

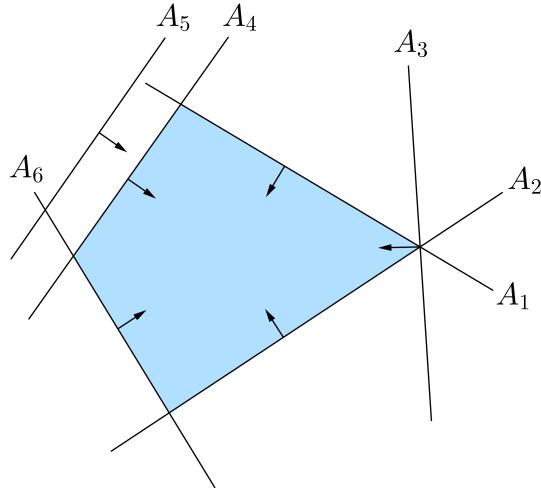


Figure 4: A Core and a Core-Determining Class

Comments: This is a stylized illustration in $\mathcal{P}(\mathcal{Y})$. Each straight line corresponds to an Artstein's inequality. The core is shaded in blue. The class of sets $\{A_1, A_2, A_4, A_6\}$ is core-determining. The sets A_1, A_2, A_4, A_6 are critical, while the sets A_3, A_5 are not.

The distribution of a random set G can be described by its *containment functional*:

$$C_G(A) \equiv P(G \subseteq A),$$

specified for all $A \in \mathbb{U}_G$. A *selection* of G is any random variable $Y : (\mathcal{U}, \mathcal{F}, P) \rightarrow (\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ that satisfies $Y(u) \in G(u)$, P -almost surely. The collection of distributions of all selections of G is called the *core*. [Artstein \(1983\)](#) showed that the core consists of all probability distributions dominating the containment functional on closed sets.

Lemma 1 (Artstein's Inequalities). *A probability distribution μ on \mathcal{Y} is the distribution of a selection of a random closed set G in \mathcal{Y} if and only if*

$$\mu(A) \geq C_G(A),$$

for all closed sets $A \subseteq \mathcal{Y}$.

Verifying Artstein's inequalities for all closed sets is a daunting task, but it usually suffices to consider smaller classes of sets. Such classes are called *core-determining*.

Definition 2.1 (Core-Determining Classes and Critical Sets). *Let G be a random closed set taking values in \mathcal{Y} with capacity functional C_G . Let \mathfrak{F} denote the class of*

all closed subsets of \mathcal{Y} . For any class of sets $\mathcal{C} \subseteq \mathfrak{F}$, denote

$$\mathcal{M}(\mathcal{C}) = \{\mu \in \mathcal{P}(\mathcal{Y}) : \mu(A) \geq C_G(A) \forall A \in \mathcal{C}\}.$$

A class \mathcal{C} is said to be core-determining if $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathfrak{F})$. A set $A \in \mathfrak{F}$ is said to be critical if $\mathcal{M}(\mathfrak{F} \setminus A) \neq \mathcal{M}(\mathfrak{F})$. Figure 4 illustrates.

Returning to identification, suppose the model postulates that $Y \in G(U, X; \theta_0)$ almost surely, for some $\theta_0 \in \Theta$. Using Artstein's theorem and the above definition, the sharp identified set for θ_0 can generally be characterized as:

$$\Theta_0 = \{\theta \in \Theta : P_{Y|X=x}(A) \geq C_{G(U, x; \theta)}(A), \text{ for all } A \in \mathcal{C}(x, \theta), \text{ a.s. } x \in \mathcal{X}\}, \quad (4)$$

where $\mathcal{C}(x, \theta) \subseteq \mathfrak{F}$ is a core-determining class for the random set $G(U, x; \theta)$ conditional on $X = x$.⁴ Equation (4) will be the working characterization for the rest of the paper.

Evidently, the smaller the core determining classes $\mathcal{C}(x, \theta)$, the simpler the above characterization. In the following section, I characterize the smallest possible core-determining class $\mathcal{C}^*(x; \theta)$, clarify how it depends on x and θ , and propose an algorithm to compute it efficiently in discrete settings.

3 The Smallest Core-Determining Class

3.1 General Case

To construct the smallest core-determining class, I identify sets $A \subseteq \mathcal{Y}$ corresponding to redundant inequalities. I proceed in three steps, borrowing the first two from [Chesher and Rosen \(2017\)](#).

First, for each $A \subseteq \mathcal{Y}$, let $\tilde{A} = \bigcup_{u \in G^-(A)} G(u)$ be the largest set from \mathbf{U}_G contained in A .⁵ Since $\tilde{A} \subseteq A$, it must be that $\mu(A) \geq \mu(\tilde{A})$, and since $G \subseteq A$ if and only if

⁴The representations via unconditional and conditional Artstein's inequalities are equivalent; see see Theorem 2.33 in [Molchanov and Molinari \(2018\)](#).

⁵Notice \tilde{A} is Borel measurable. By definition of a measurable correspondence, $G^-(A) \in \mathcal{F}$, so by Theorem 18.6 in [Aliprantis and Border \(2006\)](#) the set $\{(u, y) \in G^-(A) \times \mathcal{Y} : y \in G(u)\}$ lies in the product sigma-field $\mathcal{F} \otimes \mathcal{B}_{\mathcal{Y}}$. The above \tilde{A} is equal to the projection of this set onto \mathcal{Y} , which is measurable by definition of the product sigma-field.

$G \subseteq \tilde{A}$, it must be that $C_G(A) = C_G(\tilde{A})$. Then, given $\mu(\tilde{A}) \geq C_G(\tilde{A})$,

$$\mu(A) \geq \mu(\tilde{A}) \geq C_G(\tilde{A}) = C_G(A),$$

so A is redundant given \tilde{A} .

Second, suppose for some $A \in \mathbb{U}_G$, there are sets $A_1, A_2 \subseteq \mathbb{U}_G$ such that $A_1 \cap A_2 = \emptyset$, $A_1 \cup A_2 = A$, and $G^-(A_1 \cup A_2) = G^-(A_1) \cup G^-(A_2)$. The third condition means that $G(u) \subseteq A$ holds if and only if either $G(u) \subseteq A_1$ or $G(u) \subseteq A_2$ holds. Then, given $\mu(A_1) \geq C_G(A_1)$ and $\mu(A_2) \geq C_G(A_2)$,

$$\mu(A) = \mu(A_1) + \mu(A_2) \geq C_G(A_1) + C_G(A_2) = C_G(A),$$

so A is redundant given A_1 and A_2 .

Third, suppose that for some $A \in \mathbb{U}_G$, there are sets $A_1, A_2 \in \mathbb{U}_G$ such that $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, and $G^-(A_1) \cup G^-(A_2) = \mathcal{U}$. Here, the third condition means that for all u , either $G(u) \subseteq A_1$ or $G(u) \subseteq A_2$ holds. Then, given $\mu(A_1) \geq C_G(A_1)$ and $\mu(A_2) \geq C_G(A_2)$,

$$1 + \mu(A) = \mu(A_1) + \mu(A_2) \geq C_G(A_1) + C_G(A_2) = 1 + C_G(A),$$

so A is redundant given A_1 and A_2 . Note that the above conditions can be equivalently stated as $A_1^c \cup A_2^c = A^c$, $A_1^c \cap A_2^c = \emptyset$, and $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \emptyset$.

Therefore, any closed set $A \subseteq \mathcal{Y}$ which: (i) is not a union of elements of the support of G ; (ii) is a union of suitably disjoint “small” sets; or (iii) has a complement that is a union of suitably disjoint “small” sets, is redundant. In fact, it turns out that *any* redundant set must satisfy one of the three conditions, which allows to characterize the smallest core-determining class. In the theorem below, for any two sets $B_1, B_2 \subseteq \mathcal{U}$, the statement “ $B_1 = B_2$, P -a.s.” means $P((B_1 \setminus B_2) \cup (B_2 \setminus B_1)) = 0$.

Theorem 1. *Let $(\mathcal{U}, \mathcal{F}, P)$ be a probability space, $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ a standard Borel space, and $G : \mathcal{U} \Rightarrow \mathcal{Y}$ a random closed set. Suppose there is no $\mathcal{Y}_1, \mathcal{Y}_2$ such that $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$, $\mathcal{Y}_1 \cup \mathcal{Y}_2 = \mathcal{Y}$ and $G^{-1}(\mathcal{Y}_1) \cap G^{-1}(\mathcal{Y}_2) = \emptyset$, P -a.s., or consider each \mathcal{Y}_j separately. Then, a set $A \subseteq \mathcal{Y}$ is critical if and only if:*

1. $A \in \mathbb{U}_G$.
2. There do not exist $A_1, A_2 \in \mathbb{U}_G$ satisfying: $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \emptyset$, and

$$G^-(A_1 \cup A_2) = G^-(A_1) \cup G^-(A_2), \text{ } P\text{-a.s.}$$

3. There do not exist $A_1, A_2 \in \mathbf{U}_G$ such that: $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, and $G^-(A_1) \cup G^-(A_2) = \mathcal{U}$, $P\text{-a.s.}$

Moreover, the class \mathcal{C}^* of all critical sets is the smallest core-determining class.

The first part of the statement connects the critical inequalities to the structure of the correspondence G . The fact that the critical sets satisfy all three conditions follows from the arguments preceding the theorem. The converse is much more involved. I show that if there is a core-determining class \mathcal{C} that does not include a set A , then either condition 1 fails, or the Farkas Lemma implies that conditions 2 or 3 of the theorem must fail for some $A_1, A_2 \in \mathcal{C}$. Taking the contrapositive, if a set A satisfies all three conditions, there does not exist a core-determining class that excludes it, so that the set A must be critical. Moreover, the above implies that if A is redundant, given A_1 and A_2 , the redundancy of A_1 and A_2 does not depend on A . Therefore, all non-critical sets can be excluded from the analysis simultaneously, which implies the second part of the statement.

Since the requirements of Theorem 1 concern only the structure of the correspondence defining the random set, but not its distribution, it follows that the smallest core-determining class depends only on the support of the random set.

Corollary 1.1. *Under the assumptions of Theorem 1, the smallest core-determining class \mathcal{C}^* depends only on the support of the random set G .*

The above has immediate practical implications. Specifically, suppose for every fixed θ , the support of $G(U, x; \theta)$, conditional on $X = x$, does not depend on x . Then, in the notation of equation (4), $\mathcal{C}^*(x, \theta) = \mathcal{C}^*(\theta)$, and the sharp identified set is:

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

Alternatively, suppose the parameter space can be partitioned into equivalence classes $\Theta = \bigcup_l \Theta_l$ so that the support of $G(U, x; \theta)$, conditional on $X = x$, is the same for all $\theta \in \Theta_l$. Then, $\mathcal{C}^*(\theta, x) = \mathcal{C}_l^*(x)$ for all $\theta \in \Theta_l$, and the sharp identified set is:

$$\Theta_0 = \bigcup_l \left\{ \theta \in \Theta_l : P_{Y|X=x}(A) \geq C_{G(U,x;\theta)}(A), \text{ for all } A \in \mathcal{C}_l^*(x), x \in \mathcal{X} \right\}$$

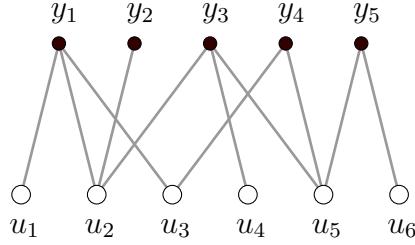


Figure 6: A bipartite graph representing a discrete random set.

If both of the above support assumptions hold, then $\mathcal{C}^*(x, \theta) = \mathcal{C}_l^*$ for all $\theta \in \Theta_l$ and $x \in \mathcal{X}$, and the sharp identified set is given by:

$$\Theta_0 = \bigcup_l \left\{ \theta \in \Theta_l : \text{ess inf}_{x \in \mathcal{X}} (P_{Y|X=x}(A) - C_{G(U,x;\theta)}(A) \geq 0), \text{ for all } A \in \mathcal{C}_l^* \right\}.$$

Detailed examples are provided in Section 3.3.

When the support of $G(U, x; \theta)$, conditional on $X = x$, is infinite, the smallest core-determining classes $\mathcal{C}^*(x, \theta)$ contain an infinite number of sets, for each x . In such settings, Theorem 1 gives a negative conclusion: if any inequality from any class $\mathcal{C}^*(x, \theta)$ is omitted, the resulting identified set will not be sharp. 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 1 can be used to guess the form of the sharp bounds, but to prove sharpness, one may have to explicitly couple the outcome with the random set. The first two examples in Section 3.3 illustrate this line of arguments.

When the support of $G(U, x; \theta)$, conditional on $X = x$, is finite, the smallest core-determining classes contain a finite number of sets, for each x , and are often substantially smaller than the power set of \mathcal{Y} . In such settings, Theorem 1 motivates an efficient algorithm for computing the smallest core-determining class numerically.

3.2 Finite Outcome Space

3.2.1 The Smallest Core-Determining Class via Subgraph Connectivity

Let $\mathcal{Y} = \{y_1, \dots, y_s\}$ denote the outcome space and $S(x, \theta) = \{G_1, \dots, G_K\}$ denote the support of $G(U, x; \theta)$, conditional on $X = x$. Partition the latent variable space as $\mathcal{U}(x, \theta) = \{u_1, \dots, u_K\}$, where $u_k = \{u \in \mathcal{U} : G(u, x; \theta) = G_k\}$, and define a

probability measure $P_{(x,\theta)}$ on $\mathcal{U}(x,\theta)$ by $P_{(x,\theta)}(u_k) = P_{U|X=x,\theta}(\{u : G(u,x;\theta) = G_k\})$. In this way, G can be viewed as a random set defined on a discrete probability space $(\mathcal{U}(x,\theta), 2^{\mathcal{U}(x,\theta)}, P_{(x,\theta)})$. Such G can be represented by an undirected bipartite graph \mathbf{B} with vertices $V(\mathbf{B}) = (\mathcal{U}, \mathcal{Y})$ and edges $E(\mathbf{B}) = \{(u, y) \in \mathcal{U} \times \mathcal{Y} : y \in G(u)\}$. The following terminology will be used. 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 connected if every vertex can be reached from any other vertex through a sequence of edges.

With the above definitions, all three conditions of Theorem 1 can be re-stated in terms of connectivity of suitable subgraphs of \mathbf{B} . For illustration, consider the graph in Figure 6. First, consider the set $A = \{y_1, y_2\}$. Since $G^-(A) = \{u_1\}$, it follows that $A \notin \mathcal{U}_G$, so by condition 1 of the theorem, A is redundant. Note that the subgraph of \mathbf{B} induced by $(A, G^-(A))$ is disconnected. Second, consider the set $A = \{y_1, y_3\}$. Then, $A_1 = \{y_1\}$ and $A_2 = \{y_3\}$ satisfy $A_1 \cup A_2 = A$ and $A_1 \cap A_2 = \emptyset$. Moreover, $G^-(A_1) = \{u_1\}$, $G^-(A_2) = \{u_4\}$, and $G^-(A) = \{u_1, u_4\}$, so that $G^-(A) = G^-(A_1) \cup G^-(A_2)$. Therefore, by condition 2 of the theorem, A is redundant. Note that as before, the subgraph of \mathbf{B} induced by $(A, G^-(A))$ is disconnected. Finally, consider the set $A = \{y_3, y_4\}$. Then, $A_1 = \{y_1, y_2, y_3, y_4\}$ and $A_2 = \{y_3, y_4, y_5\}$ satisfy $A^c = A_1^c \cup A_2^c$ and $A_1^c \cap A_2^c = \emptyset$. Moreover, $G^{-1}(A_1^c) = \{u_5, u_6\}$ and $G^{-1}(A_2^c) = \{u_1, u_2, u_3\}$, so that $G^{-1}(A_1^c) \cap G^{-1}(A_2^c) = \emptyset$. Therefore, by condition 3 of the theorem, A is redundant. Note that in this case, the subgraph of \mathbf{B} induced by $(A^c, G^{-1}(A^c))$ is disconnected.

Therefore, for any set A that is redundant according to Theorem 1, at least one of the subgraphs of \mathbf{B} , induced by $(A, G^-(A))$ or $(A^c, G^{-1}(A^c))$, must be disconnected. Conversely, one can verify that any set A , for which one of the induced subgraphs is disconnected, must fail to satisfy at least one of the conditions of the theorem. Thus, in discrete settings, Theorem 1 can be restated as follows.

Theorem 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 random set with a bipartite graph \mathbf{B} . Suppose, without loss of generality, that $P(u_k) > 0$ for all $u_k \in \mathcal{U}$. Suppose that \mathbf{B} is connected, or consider each connected component separately. A set $A \subseteq \mathcal{Y}$ is critical if and only if:*

1. *The subgraph of \mathbf{B} induced by $(A, G^-(A))$ is connected.*
2. *The subgraph of \mathbf{B} induced by $(A^c, G^{-1}(A^c))$ is connected.*

Moreover, the class \mathcal{C}^* of all critical sets is the smallest core-determining class.

This result has two immediate practical implications. First, it explicitly shows that the smallest core-determining class depends only on the structure of the bipartite graph, i.e, the support of the random set.⁶ In the notation of Equation (4), if the support of $G(U, x; \theta)$, conditional on $X = x$, does not change with x or θ within some equivalence classes and $P_{(x,\theta)}(u) > 0$ for all $u \in \mathcal{U}(x, \theta)$, the same moment inequalities are non-redundant across all such x and θ . The examples in Section 3.3 demonstrate that these conditions are generically satisfied in discrete-outcome models. Second, Theorem 2 suggests that the smallest core-determining class can be computed by checking connectivity of certain subgraphs of \mathbf{B} . Importantly, it motivates an algorithm that looks (almost) exclusively at critical inequalities, thus avoiding the computational bottleneck of checking all $2^{|\mathcal{Y}|} - 2$ candidate inequalities for redundancy. The details are discussed in Section 4. For a fixed size of the outcome space, $|\mathcal{Y}|$, the size of the smallest core-determining class may be drastically 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 core-determining class ex ante, useful benchmarks can be provided. Appendix A.4 gives a detailed discussion.

3.3 Discussion and Applications

This section shows Theorems 1 and 2 in action, using the four examples introduced in Section 2.1, a dynamic monopoly entry model from [Berry and Compiani \(2020\)](#), and a network formation model from [Gualdani \(2021\)](#).

The first two examples deal with settings where the outcome variable is continuous. In such settings, the smallest core-determining class contains an infinite number of sets, making the sharp identified set of the parameter vector θ_0 intractable. However, certain functionals of interest $\phi(\theta_0)$ still admit tractable sharp bounds, which can be easily deduced from Theorem 1 but may be hard to obtain using other methods. To prove sharpness, it is typically easier to explicitly construct the distributions

⁶The most general implication requires an extra definition. Graphs \mathbf{B} and \mathbf{B}' are said to be f -isomorphic if there is a bijection $f : V(\mathbf{B}) \rightarrow V(\mathbf{B}')$ so that $(u, y) \in E(\mathbf{B})$ if and only if $(f(u), f(v)) \in E(\mathbf{B}')$. Consider two random sets $G : (\mathcal{U}, 2^{\mathcal{U}}, P) \rightrightarrows \mathcal{Y}$ and $G' : (\mathcal{U}', 2^{\mathcal{U}'}, P') \rightrightarrows \mathcal{Y}$ with the graphs \mathbf{B} and \mathbf{B}' . If the graphs are f -isomorphic and $P(\{u\}) > 0$ whenever $P'(\{f(u)\}) > 0$, then G and G' have the same core-determining classes.

that attain the bounds than to verify all of the inequalities from Theorem 1.

Example 3 (Continued). Consider the model with interval-observed outcomes, where Y^* denotes the latent outcome variable satisfying $Y^* \in G(Y_L, Y_U) = [Y_L, Y_U]$, and $X \in \mathcal{X}$ is a vector of covariates. For simplicity, suppose X is discrete, and $\mathcal{Y} = [\underline{y}, \bar{y}]$. Additionally, assume that the joint distribution of (Y_L, Y_U) , conditional on $X = x$, satisfies $P(\underline{\kappa}(x) \leq Y_U - Y_L \leq \bar{\kappa}(x) | X = x) = 1$, for some known functions $\underline{\kappa}(x)$ and $\bar{\kappa}(x)$. The primitive parameter of interest is the joint distribution $\theta_0 = P_{Y^*|X}$.

The class \mathcal{U}_G contains all sets that can be expressed as a finite or countable union of disjoint intervals included in $[\underline{y}, \bar{y}]$, where each interval has length of at least $\underline{\kappa}(x)$. Further, for a union $A = A_1 \cup A_2 = [a_1, b_1] \cup [a_2, b_2]$ with $b_j - a_j \geq \underline{\kappa}(x)$ and $a_2 > b_1$, one has $A_1 \cap A_2 = \emptyset$ and $G^-(A_1) \cap G^-(A_2) = \emptyset$, P -a.s. Therefore, this set fails to satisfy condition (2) of Theorem 1. A similar argument applies to any other collection of disjoint intervals. Therefore, the only sets that satisfy conditions (1) and (2) of the theorem are contiguous intervals with length at least $\underline{\kappa}(x)$. This class includes boundary intervals of the form $[\underline{y}, b]$ or $[a, \bar{y}]$ and interior intervals of the form $[a, b]$. Consider an interior interval $A = [a, b]$ with $b - a > \bar{\kappa}(x)$. Note that the sets $A_1 = [\underline{y}, b]$ and $A_2 = [a, \bar{y}]$ satisfy $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, and $G^-(A) \cap G^-(A_2) = \mathcal{U}$, P -a.s. Therefore, the set A fails to satisfy condition (3) of the theorem. Note that, even though the boundary sets may have length exceeding $\bar{\kappa}(x)$, they still satisfy condition (3) of the theorem. Therefore, the sharp identified set for θ_0 is characterized by the inequalities:

$$P(Y^* \in A | X = x) \geq P([Y_L, Y_U] \subseteq A | X = x),$$

for all A in

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

for all $x \in \mathcal{X}$. If $\underline{\kappa}(x)$ or $\bar{\kappa}(x)$ are constant for all $x \in \mathcal{X}'$, for some $\mathcal{X}' \subseteq \mathcal{X}$, the corresponding inequalities can be intersected. Importantly, Theorem 1 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:

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

The upper and lower bounds are sharp in the sense of the 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

$$F_{Y^*|X=x}(b) - F_{Y^*|X=x}(a) \geq P(Y_L \geq a, Y_U \leq b | X = x) \quad (5)$$

for any a, b such that $\underline{\kappa}(x) \leq b - a \leq \bar{\kappa}(x)$. 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:

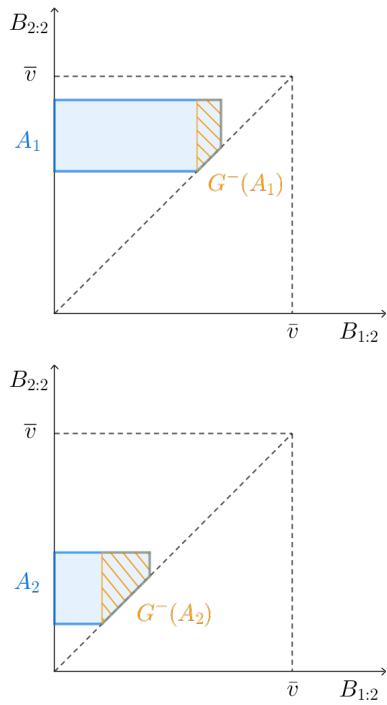
$$\phi(\theta_0) \in [\max\{0, q_{Y_L|X=x}(\tau_1) - q_{Y_U|X=x}(\tau_2)\}, q_{Y_U|X=x}(\tau_1) - q_{Y_L|X=x}(\tau_2)].$$

However, the upper bound may not be sharp due to (5) being violated at $a = q_{Y^*|X=x}(\tau_2)$, $b = q_{Y^*|X=x}(\tau_1)$. Instead, it follows that the sharp upper bound is:

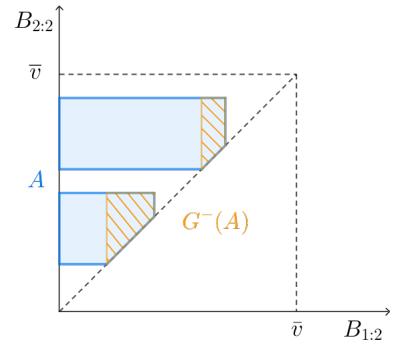
$$\max\{b - a | a \geq q_{Y_L|X=x}(\tau_2), b \leq q_{Y_U|X=x}(\tau_1), \tau_1 - \tau_2 \geq P(Y_L \geq a, Y_U \leq b | X = x)\}.$$

Sharp bounds on other functionals of interest can be constructed similarly; see [Molinari \(2020\)](#) for related results. ■

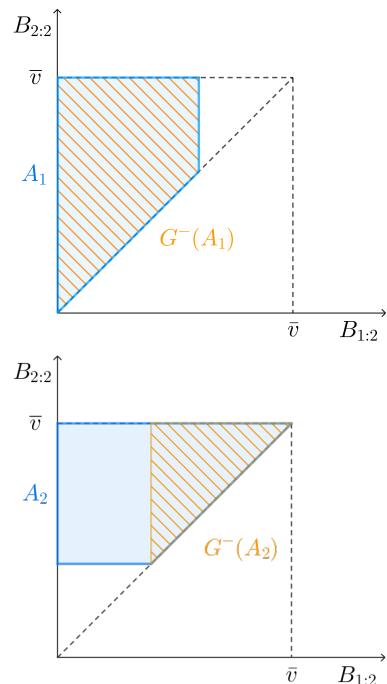
Example 4 (Continued). For simplicity, suppose $N = 2$, or only the top two bids are observed. Recall that the primitive parameter of interest is the joint distribution of ordered valuations, $\theta_0 = F$. The set \mathbf{U}_G consists of all lower sets, $A_1 = \{(v_1, v_2) \in S : v_2 \leq \kappa_1(v_1)\}$, where $\kappa_1 : [0, \bar{v}] \rightarrow [0, \bar{v}]$ is any decreasing function, all sets of the form $A_2 = \{(v_1, v_2) \in S : v_1 \leq a, v_2 \in [b, c]\}$, all sets of the form $A_1 \cap A_2$, and all unions of the resulting family of sets. Figure 7 presents some examples of redundant sets $A \in \mathbf{U}_G$ that fail to satisfy conditions (2) or (3) of the theorem. In the upper



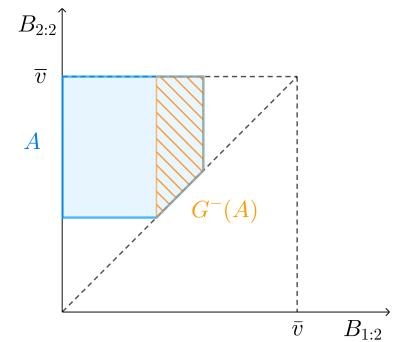
(a) “Small” sets.



(b) Redundant union of “small” sets.



(c) “Large” sets.



(d) Redundant intersection of “large” sets.

Figure 7: Application of Theorem 1 to the English auction model in Example 4.

panel, $A_1 \cap A_2 = \emptyset$, $A_1 \cup A_2 = A$, and $G^-(A_1 \cup A_2) = G^-(A_1) \cup G^-(A_2)$. In the lower panel, $A_1 \cap A_2 = A$, $A_1 \cup A_2 = \mathcal{Y}$, and $G^-(A) = G^-(A_1) \cap G^-(A_2)$. All other sets in \mathbb{U}_G are critical. Therefore, the smallest core-determining class is infinite and involves rather complicated sets, so the sharp identified set for the joint distribution of valuations is intractable.

However, the joint distribution is typically of interest only to the extent that it allows to calculate counterfactual quantities of interest. For example, [Aradillas-López, Gandhi, and Quint \(2013\)](#) note that the expected profit and bidders surplus in English auctions with equilibrium play 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:

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

where \mathcal{F} denotes the set of all joint distributions satisfying the desired assumptions on the information structure. To make progress, following [Aradillas-López, Gandhi, and Quint \(2013\)](#), suppose that the valuations are positively dependent in the sense that the probability $P(V_i \leq v | \#\{j \neq i : V_j \leq v\} = k)$ is non-decreasing in k , for each $i = 1, \dots, N$. The authors show that, under the above assumption, $F_{N:N} \in [F_{N-1:N}, \phi_{N-1:N}(F_{N-1:N})^N]$, where $\phi_{N-1:N} : [0, 1] \rightarrow [0, 1]$ is a known strictly increasing function that maps the distribution of the second-largest order statistic of an i.i.d. sample of size N to the parent distribution.

The Artstein's inequality with the set $A = S \cap [0, v] \times [0, \bar{v}]$ implies the upper bound $F_{N-1:N}(v) \leq G_{N-1:N}(v)$, and the set $A = S \cap [0, \bar{v}] \times [v, \bar{v}]$ implies the lower bound $F_{N-1:N}(v) \geq G_{N:N}(v)$. Additionally, the set $A = S \cap [0, v] \times [0, v]$ implies $F_{N:N}(v) \leq G_{N:N}(v)$. Combining these inequalities with the imposed positive-dependence assumption on F yields the following bounds:

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

By constructing a suitable joint distribution $F \in \mathcal{F}$, it is possible to show that both upper bounds and both lower bounds can be attained simultaneously, so the above 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 $A = S \cap [a, \bar{v}] \times [0, b]$ for $b > a$ corresponds to the inequality $F_{N-1:N}(b) - F_{N-1:N}(a) \geq P(B_{N-1:N} \geq a, B_{N:N} \leq b)$, which rules out the CDF-s that do not increase sufficiently between a and b . Paralleling the preceding example, this has implications for studying differences in expected profit under different reserve prices. The details are left for further research. ■

The remaining examples illustrate applications of Theorem 2 in discrete settings. Typically, the smallest core-determining class contains much fewer sets than the power set of the outcome space. In many settings, the difference is of several orders of magnitude, and selecting the “correct” set of inequalities keeps the analysis tractable while still delivering sharp identified sets. However, in several commonly studied models, even the smallest core-determining class is too large to be used in practice, so additional inequality selection is required. In this context, Examples 1 and 6 discuss imposing additional assumptions on the structure of the model’s correspondence, or the equilibrium selection mechanism, to simplify the analysis. Importantly, the resulting identified set Θ'_0 contains the original sharp identified set Θ_0 , and the two sets coincide if the extra assumptions are satisfied.

Example 1 (Continued). First, consider the entry game with substitutes, i.e., $\delta_j < 0$ for all j , in which the firms compete against each other upon entering the market. In this case, 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: $\mathcal{Y} = \bigcup_{n=0}^N \mathcal{Y}_n$. This property dramatically reduces the size of core-determining class because all sets of the form $A = \bigcup_{n=0}^N A_n$, where $A_n \subseteq \mathcal{Y}_n$, are redundant.⁷ For $N = 2$, the partition of the space of latent variables is illustrated in Figure 1. While the regions in the partition and the corresponding probabilities change with the values of $\theta = (\alpha_j, \delta_j)_{j=1}^n$, the bipartite graph remains the same as long as all $\delta_j < 0$. Therefore, the core-determining class only needs to be computed once. Table 1 summarizes the results for $N = 2, \dots, 6$. Column “CR” corresponds to the result of Chesher and Rosen (2017), and is computed by checking only condition 1 from Theorem 2. While the smallest core-determining class is substantially smaller than the power set of the outcome space, it quickly becomes intractable.

⁷This fact follows from Theorem 2, or, alternatively, Theorem 3 from Chesher and Rosen (2017), or Theorem 2.23 from Molchanov and Molinari (2018).

Next, consider the entry game with complementarities, i.e., $\delta_j > 0$, which corresponds to 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 $\theta = (\alpha_j, \delta_j)_{j=1}^n$, the bipartite graph stays the same as long as all $\delta_j > 0$, so, the smallest core-determining class only needs to be computed once. Table 1 summarizes the results for $N = 2, \dots, 6$. As before, even the smallest core-determining class quickly becomes intractable.

To simplify the analysis, one may restrict firm heterogeneity by assuming there are several types of firms, and all firms within each type are identical, including the unobserved cost shifters.⁸ For example, suppose there are two types of firms, $t \in \{1, 2\}$, and the profit functions depend only on the number of entrants of each type:

$$\begin{aligned}\pi_i^1(Y) &= \alpha_1 + \alpha_2 \cdot (N(Y) - 1) + \varepsilon_1 \\ \pi_i^2(Y) &= \beta_1 + \beta_2 \cdot N_1(Y) + \beta_3 \cdot (N_2(Y) - 1) + \varepsilon_2.\end{aligned}$$

Here, $Y \in \{0, 1\}^N$, the subscript i refers to the firm, the superscripts index firm types, and $N_t(Y)$ is the number of entrants of type t . Moreover, suppose that $\alpha_1, \beta_2, \beta_3 < 0$ and $\beta_3 \geq \beta_2$. With $\beta_3 = \beta_2$, this is a direct simplification of the fully heterogeneous model discussed above. With $\beta_3 > \beta_2$, the firms compete in an asymmetric way (e.g. type-1 firms are large and type-2 firms are small).

From the identification perspective, the 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. Therefore, in the absence of assumptions about equilibrium selection, it is without loss of information to re-define the outcome space in terms of the numbers of entrants of each type. That is, $Y = (N_1, N_2)$, and letting \bar{N}_j denote the number of potential entrants of type $j = 1, 2$, the outcome space is $\mathcal{Y} = \{0, 1, \dots, \bar{N}_1\} \times \{0, 1, \dots, \bar{N}_2\}$. Since the outcome space has a much smaller card-

⁸An version of this model with only one type leads back to [Bresnahan and Reiss \(1991\)](#). The model with two types, discussed here, was proposed by [Berry and Tamer \(2006\)](#) and studied in [Galichon and Henry \(2011\)](#) and [Luo and Wang \(2018\)](#).

Setting:		$\delta_j < 0$		$\delta_j > 0$	
N	Total	CR	Smallest	CR	Smallest
2	14	5	5	5	5
3	254	16	16	86	14
4	65534	95	95	43309	23770
5	$\sim 10^9$	2110	2110	—	—
6	$\sim 10^{19}$	$\sim 10^6$	$\sim 10^6$	—	—

Table 1: Sizes of core-determining classes in the entry game in Example 1.

(\bar{N}_1, \bar{N}_2)	Total	CR	Smallest
(2, 2)	$2^9 - 2$	33	14
(3, 3)	$2^{16} - 2$	365	25
(4, 4)	$2^{25} - 2$	33975	50
(5, 5)	$2^{36} - 2$	-	97
(6, 6)	$2^{49} - 2$	-	152
(2, 7)	$2^{24} - 2$	351	57

Table 2: Sizes of core-determining classes in the entry game with types and $\beta_3 > \beta_2$ in Example 1.

nality and leads to simpler bipartite graphs, the size of the smallest core-determining class is much smaller. Table 2 presents the results for different compositions of firms.

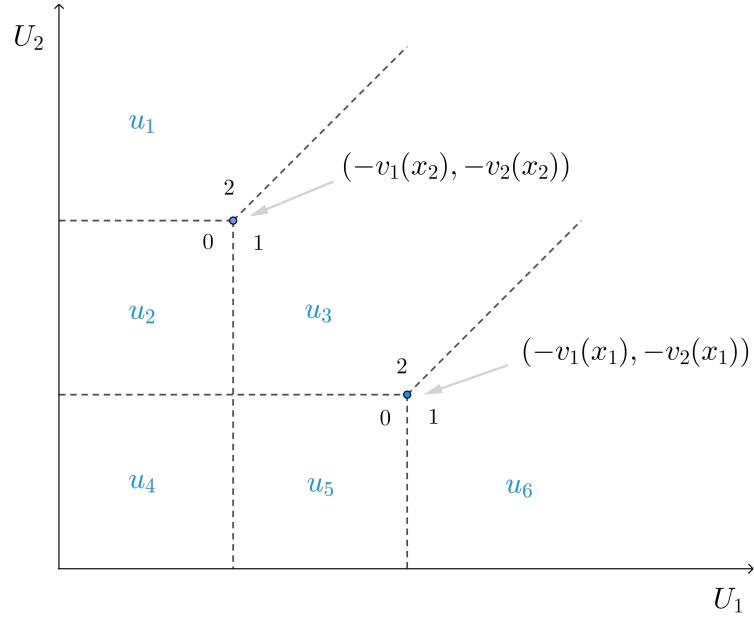
Notably, the resulting set of Artstein's inequalities is a subset of the inequalities in fully heterogeneous model. Therefore, the resulting sharp identified set Θ'_0 will always contain the sharp identified set Θ_0 in the fully heterogeneous model. Moreover, if the type-heterogeneity assumption is satisfied by the distribution of the data, the two sets will coincide. In this way, the type-heterogeneity assumption can serve as a formal basis for inequality selection. ■

K	2	3	4	5	6	7	8	9	10
Total	62	510	4094	32766	262142	$2^{21} - 2$	$2^{24} - 2$	$2^{27} - 2$	$2^{30} - 2$
Smallest	12	33	82	188	406	842	1703	3397	6733

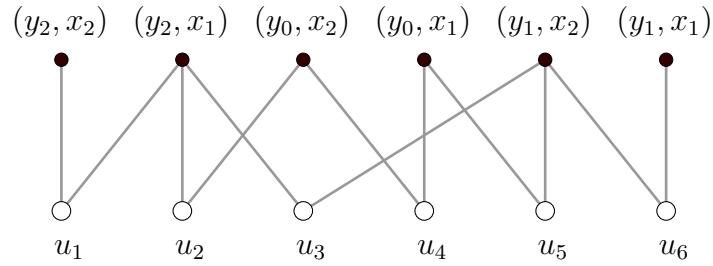
Table 3: Core-determining classes in the discrete choice model from Example 2.

T	2	3	4	5	6	7	8	9	10
Total	30	65534	$\sim 10^9$	$\sim 10^{19}$	$\sim 10^{38}$	$\sim 10^{77}$	$\sim 10^{154}$	$\sim 10^{308}$	$\sim 10^{616}$
Smallest	10	22	46	94	190	382	766	1534	3070

Table 4: Core-determining classes in the dynamic entry model from Example 5.



(a) Partition of the space of latent variables with $x \in \{x_1, x_2\}$.



(b) Bipartite graph with $x \in \{x_1, x_2\}$.

Figure 8: Illustrations for the discrete choice model in Example 2.

Example 2 (Continued). Consider the model with $J = 2$ and $\mathcal{X} = \{x_1, x_2\}$. The first step is to construct an appropriate partition of the space of latent variables. Elements of the partition take the form $\bigcap_{i,j} G(y_i, x_j; \theta)$ for some i, j . This is precisely the Minimal Relevant Partition in the sense of [Tebaldi, Torgovitsky, and Yang \(2019\)](#). Figure 8a presents one possible configuration. Numbers 0, 1, and 2 next to each point indicate the regions of latent variables for which the optimal choice is $y \in \{y_0, y_1, y_2\}$, for a given $x \in \{x_1, x_2\}$.

One can proceed by working conditional on Z only, as [Chesher, Rosen, and Smolinski \(2013\)](#), or conditional on both X and Z , as [Tebaldi, Torgovitsky, and Yang \(2019\)](#) maintaining the independence assumption. These approaches are, in theory, equivalent. Indeed, Lemma 1 implies that to characterize the set of distributions of all selections of a random set that satisfy certain restrictions, one simply has to intersect the core of G , described by Arstein's inequalities, with the set of all distributions satisfying the desired restrictions. Moreover, Theorem 2.33 in [Molchanov and Molinari \(2018\)](#) shows that conditional and unconditional Artstein's inequalities deliver the same characterization. The proof of this result is not affected by any additional constraints on the distributions of selections. Therefore, working conditional on X is equivalent to working conditional on (X, Z) and maintaining the constraints $F_{U|Z=z}(u) = F_U(u)$ for all $u \in \mathcal{U}$ and $z \in \mathcal{Z}$. Consequently, the result of [Tebaldi, Torgovitsky, and Yang \(2019\)](#) can also be obtained as follows. Conditional on (X, Z) , the smallest core-determining class contains of $J+1$ sets $\{G(y_0, x; \theta), \dots, G(y_J, x; \theta)\}$ and the corresponding conditional moment inequalities turn into equalities because $\{Y = y_j \mid X = x\} \iff \{U \in G(y_j, x; \theta) \mid X = x\}$. Combined with the constraints $F_{U|Z=z}(u) = F_U(u)$ for all $u \in \mathcal{U}$ and $z \in \mathcal{Z}$, this leads to the characterization in the aforementioned paper. If the functional of interest depends only on the probabilities of cells in the partition, and both X and Z are discrete, the approach of [Tebaldi, Torgovitsky, and Yang \(2019\)](#) brings substantial computational advantages. Otherwise, it is not applicable.

The discussion below applies conditional on Z only, following [Chesher, Rosen, and Smolinski \(2013\)](#). Figure 8b presents the corresponding bipartite graph. In this example, although the partition is different depending on the relative positions of points $(v_1(x_k), v_2(x_k))$, the structure of the corresponding bipartite graph remains the same up to a vertex isomorphism. If $x \in \{x_1, \dots, x_K\}$, there will be $(J+1)K$ vertices in the \mathcal{Y} part, and $(K+1)(K+2)/2$ vertices in the \mathcal{U} part. The total number

of non-trivial subsets of the outcome space is therefore $2^{(J+1)K} - 2$. Table 3 lists the size of the smallest core-determining class for $K = 2, \dots, 8$. In this example, while it is possible to reorder the elements of \mathcal{U} and \mathcal{Y} to preserve marginal monotonicity, the smallest core-determining class still grows exponentially in the size of \mathcal{Y} . ■

The following example is a dynamic model with endogenous states adapted from [Berry and Compiani \(2020\)](#).

Example 5 (Dynamic Monopoly Entry). In time period $t = 1, \dots, T$, a firm decides to stay out of or enter the market, $a_t \in \{0, 1\}$. The per-period profit is:

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

where $X_t \in \{0, 1\}$ indicates if the firm was active in period $t - 1$, ε_t is the variation in fixed costs, observed by the firm but not the econometrician, and $(\bar{\pi}, \gamma)$ are fixed profit and sunk costs of entering the market correspondingly. The parameters $\bar{\pi}$ and γ may depend on exogenous characteristics, which are assumed away for simplicity. For concreteness, suppose that $\varepsilon_t = \rho \varepsilon_{t-1} + \sqrt{1-\rho^2} v_t$ for some $\rho < 1$ and v_t i.i.d. $N(0, 1)$, although this does not play a role in the discussion below. The Bellman equation for the firm's problem is:

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

Under standard conditions, there is a unique stationary solution:

$$A_t = \mathbf{1}(U_t \leq \tau(X_t)),$$

where U_t , which is marginally $U[0, 1]$ distributed, is a quantile transformation of ε_t , and τ is an increasing function of X_t known up to the parameters $(\bar{\pi}, \gamma, \rho)$. If X_1 was exogenous, one could proceed with the analysis conditional on X_1 . However, it is not, and the model does not specify how it is determined. Thus, given $U = (u_1, \dots, u_T)$ and $\theta = (\bar{\pi}, \gamma, \rho)$, the model produces a set of predictions for $Y = (X_1, A_1, \dots, A_T)$:

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

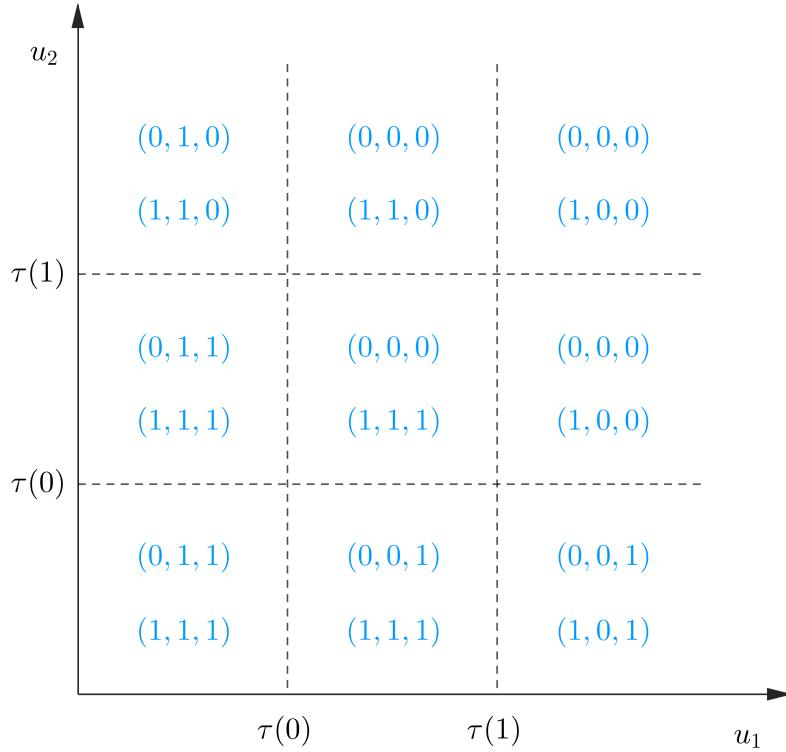


Figure 9: Set-valued prediction in the dynamic model from Example 5 with $T = 2$.

Figure 9 illustrates. The bipartite graph \mathbf{B} corresponding to the random set G has an implicit monotone structure. As a result, the size of the core-determining class scales very slowly with the size of the outcome space; see Appendix A.4 for related discussion. Table 4 summarizes the results for $T \in \{1, \dots, 10\}$.

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

Example 6 (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 $Y^k = (Y_1^k, \dots, Y_N^k) \in \mathcal{Y}^k$ 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 $\varepsilon^1, \dots, \varepsilon^N$ and the corresponding selection mechanisms are mutually independent.

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

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

$$\Theta'_0 = \{\theta \in \Theta : P(Y^k \in A^k) \geq P(G^k \subseteq A^k) \forall A^k \subseteq \mathcal{Y}^k, \forall k\}$$

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 find a tractable characterization of Θ'_0 , Theorem 2 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 $\sim 2^{64}$ inequalities in total and only 144 in the smallest class. For $N = 5$, there are $\sim 2^{1024}$ inequalities in total, and 95080 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 Computation and Relation to Other Methods

4.1 Computing the Smallest CDC in Discrete Settings

To describe the proposed algorithm, I use the following terminology. A set $A \subseteq \mathcal{Y}$ is called *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. Then, by Theorem 2, a set $A \subseteq \mathcal{Y}$ is critical if and only if it is both self- and complement-connected, and the smallest core-determining class consists of all critical sets. Additionally, a set C is called a *minimal critical superset* of A if there is no critical set \tilde{C} such that $A \subset \tilde{C} \subset C$.

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 $F(\mathcal{Y}) = \emptyset$. 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 1 (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 $Q = \{A \cup G(u) : u \in G^{-1}(A) \setminus G^-(A)\}$.
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 $l = 1, \dots, L$.
 - Collect all sets of the form $C \cup \bigcup_{j \neq l} \mathcal{Y}_j$ for $l = 1, \dots, L$ into $P(C)$.
3. Return $\bigcup_{C \in Q} P(C)$.

The construction is motivated by two observations: First, since any critical superset must be self-connected, it suffices to consider the sets in Q ; Second, if for some $C \in Q$, the subgraph of \mathbf{B} induced by $(C^c, G^{-1}(C^c))$ breaks down into several disconnected components, any minimal critical superset must contain all but one of the \mathcal{Y}_l parts of these components, because all other configurations cannot be complement-connected. The smallest core-determining class can be computed as follows.

Algorithm 2 (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 $R_k = \emptyset$.
 - 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 from Algorithm 1. Iterate on $F(\cdot)$ starting from \mathcal{C}_k until convergence and collect all sets along the way into R_k .
3. Return $\bigcup_{k=1}^K R_k \setminus \mathcal{Y}$.

Since at every iteration, except possibly the first one, Algorithm 1 is only applied to critical sets, the complexity of Algorithm 2 is proportional to the number of critical sets times the cost of decomposing subgraphs of \mathbf{B} into connected components. That is, the worst-case complexity is of order $|CDC| \times (|\mathcal{Y}| + |\mathcal{U}| + |\mathcal{E}|)$, where $|CDC|$ is the size of the smallest core-determining class, and $|\mathcal{Y}| + |\mathcal{U}|$ and $|\mathcal{E}|$ are the numbers of vertices and edges in the graph \mathbf{B} correspondingly.

4.2 The Master Algorithm

The identified set Θ_0 in Equation (4) can be obtained as follows.

Algorithm 3 (Sharp Identified Set).

1. Discretize Y and X if appropriate, and denote the resulting spaces by $\mathcal{Y} = \{y_1, \dots, y_S\}$ and $\mathcal{X} = \{x_1, \dots, x_R\}$. **Note:** if either Y or X is continuous, this procedure will produce an outer region of Θ_0 .⁹
2. Fix some $x \in \mathcal{X}$. Partition the parameter space, $\Theta = \bigcup_{l=1}^L \Theta_l(x)$, so that the support of $G(U, x; \theta)$, conditional on $X = x$, does not change with θ within each $\Theta_l(x)$. **Note:** this can be done either analytically or numerically, e.g., using Algorithm 3 from [Gu, Russell, and Stringham \(2022\)](#).
3. Fix an arbitrary $\theta \in \Theta_l(x)$. Partition the space of latent variables 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 $k = 1, \dots, K$.

⁹The finer the discretization, the closer the result will be to Θ_0 . To be precise, the preceding claim holds if the discretizations are nested in a sense that any set from the coarser discretization is a union of sets from the finer discretization.

4. Construct the bipartite graph \mathbf{B} with vertices $\mathcal{U}_{(x,\theta)}$ and \mathcal{Y} , and edges (u_k, y_s) whenever $y_s \in G(u_k, x; \theta)$, for all $s = 1, \dots, S$ and $k = 1, \dots, K$.
5. Apply Algorithm 2 to compute the smallest core-determining class $\mathcal{C}_l^*(x)$.
6. Repeating steps 2 through 5, compute the classes $\mathcal{C}_l^*(x)$ for all $x \in \mathcal{X}$ and $l = 1, \dots, L$, and construct Θ_0 .

■

4.3 Relation to Other Existing Approaches

There are several closely related approaches to sharp identification in the class of models studied in this paper. The common idea is to find a tractable characterization for the set of conditional distributions of the outcome, given covariates, $\mathcal{P}(x; \theta)$.

The first approach is based on Artstein’s inequalities, and it was explored in [Beresteanu and Molinari \(2008\)](#), [Galichon and Henry \(2011\)](#), [Chesher, Rosen, and Smolinski \(2013\)](#), [Chesher and Rosen \(2017, 2020\)](#), [Molinari \(2020\)](#), and [Luo and Wang \(2018\)](#). Here, the set $\mathcal{P}(x; \theta)$ is expressed as the core of the random set $G(U, x; \theta)$, conditional on $X = x$. This paper directly extends the results of [Galichon and Henry \(2011\)](#) and [Chesher and Rosen \(2017, 2020\)](#), as has already been discussed. [Luo and Wang \(2018\)](#) propose a different characterization of the smallest core-determining class in discrete settings. Upon close inspection, the set of inequalities resulting from Theorem 2 is the same as in that paper. However, the new characterization is simpler and substantially accelerates computation. In many scenarios, Algorithm 2 is exponentially faster than Algorithm 1 of [Luo and Wang \(2018\)](#), as they cannot get around the bottleneck of explicitly checking all possible inequalities for redundancy. Moreover, further substantial improvements over Algorithm 2 do not appear possible: the goal is to find all critical sets, and the complexity of the algorithm is proportional to the number of such sets.

The second approach, proposed by [Beresteanu, Molchanov, and Molinari \(2011\)](#), represents the set $\mathcal{P}(x; \theta)$ as the Aumann expectation of a suitably defined random set. The Aumann expectation, denoted $\mathbf{E}[G(U, X; \theta)|X]$, is defined as the set of conditional expectations of all of its measurable selections, so $\mathbb{E}[Y|X] \in \mathbf{E}[G(U, X; \theta)|X]$ almost surely if and only if Y is a measurable selection of $G(U, X; \theta)$. In non-atomic

probability spaces, the Aumann expectation is always convex, so it can be characterized via the support function, $h_{\mathbf{E}[G|X]}(t) = \sup_{a \in \mathbf{E}[G|X]} a^T t$, defined on the unit ball, $t \in B_1$. While the Aumann expectaion itself may be hard to compute in practice, one can show that $h_{\mathbf{E}[G|X]}(u) = \mathbb{E}[h_G(u)|X]$ holds almost-surely. Thus, the sharp identified set is given by $\{\theta \in \Theta : t^T \mathbb{E}[Y|X = x] \leq \mathbb{E}[h_{G(U,x;\theta)}(t)|X = x] \forall t \in B_1\}$, which can be reduced to a convex minimization problem. This approach is advantageous in models where parameters of interest can be directly related to expectations of random sets. As discussed in [Beresteanu, Molchanov, and Molinari \(2011\)](#) and [Molchanov and Molinari \(2018\)](#), in discrete outcome models, this approach is equivalent to verifying all Artstein's inequalities and does not lead to computational advantages. A somewhat related approach appears in [Mbakop \(2023\)](#). The paper studies panel discrete-outcome models and characterizes the set $\mathcal{P}(x; \theta)$ by solving multiple-objective linear programs. Although a similar approach can be applied to some of the models studied here, Algorithm 2 scales more favorably to larger outcome spaces.

The third approach, proposed by [Tebaldi, Torgovitsky, and Yang \(2019\)](#) and [Gu, Russell, and Stringham \(2022\)](#), characterizes sharp identified sets for a class of counterfactuals in discrete-outcome models using linear programming. The former paper was discussed in detail in Example 2, so here I focus on the latter. 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^*, X^*) , denoted $\psi(P_{Y^*|X^*})$. The counterfactual set of predictions G^* is assumed to be “coarser” than the factual set G in the following sense: there must exist a partition $\{u_1, \dots, u_L\}$ of the latent variable space \mathcal{U} such that knowing probabilities of cells u_k , conditional on $X = x$, suffices to bound the counterfactual of interest. Paralleling [Tebaldi, Torgovitsky, and Yang \(2019\)](#), this partition is called the minimal relevant partition. The authors show that $Y \in G(U, X; \theta)$ and $Y^* \in G^*(U, X^*; \theta)$ holds jointly if and only if there exists a joint selection mechanism $\pi_x(y, y^*, u_l)$ consistent with the model. Here, $\pi_x(y, y^*, u_l)$ represents the probability that an factual outcome y is chosen from the set G and a counterfactual outcome y^* is chosen from the set G^* , given $u \in u_l$, conditional on x . This allows to express sharp bounds on the counterfactual of interest using two linear programs. The choice vector $\{\pi_x(y, y^*, u_l)\}$ is of dimension $d = |\mathcal{X}||\mathcal{Y}|^2 L$, and there are $p = |\mathcal{X}|(|\mathcal{Y}| + 2)$ constraints that ensure that

it is a valid selection mechanism, and $q = |\mathcal{X}||\mathcal{Y}|^2L$ non-negativity constraints. The approach developed in this paper can sometimes be used to simplify the resulting linear programs. Specifically, since the bounds on the counterfactual of interest depend only on the probabilities of cells u_l , conditional on $X = x$, one can treat them directly as unknown parameters, denoted $\mu(u_l, x)$. These cells are typically finer than the partition of the latent variable space described in Section 3.2, denoted here by $\{\mathcal{U}_1, \dots, \mathcal{U}_K\}$, so each $\mu(\mathcal{U}_k, x)$ will be a sum of several $\mu(u_l, x)$. In this setting, Artstein’s inequalities will take the form $P(Y \in A | X = x) \geq \sum_{k \in G^-(A)} \mu(\mathcal{U}_k, x)$ for all $A \in \mathcal{C}^*(x)$. Assuming $|\mathcal{C}^*(x)|$ does not change with x , for simplicity, this approach leads to a linear program with the choice vector $\{\mu(u_l, x)\}$ of dimension $d = |\mathcal{X}|L$, $p = |\mathcal{X}|K$ equality constraints linking the two partitions, and $q = |\mathcal{X}|(|\mathcal{C}^*(x)| + L)$ inequality constraints including the Artstein’s inequalities and non-negativity constraints. If $|\mathcal{Y}|$ is moderate or large, and $|\mathcal{C}^*(x)|$ is manageable, such linear program will be easier to solve.

5 Numerical Illustration: The Importance of Selecting Inequalities

In this section, using a small scale simulation, I demonstrate the importance of using the “correct” set of inequalities for identification. I consider the market entry model from Example 1 with $N = 3$ players and strategic complementarities, i.e., $\delta_j > 0$. Parameters of the simulation are chosen as follows. The true values are $\alpha_j = 1$ and $\delta_j = 1.75$ for all three firms. The unobservables ε_j are i.i.d. across firms and follow a Normal distribution with parameters $\mu = 0, \sigma = 5$. Within the regions of multiplicity, each of the equilibria is selected with equal probability. The sample size is $n = 50,000$. To construct the identified sets, I perform a grid search with radius 1.5 around the true values with step size 0.01.

In total, there are 254 non-trivial subsets of the outcome space, each corresponding to an Artstein’s inequality. I compare two approaches for constructing identified sets: (1) use the 14 non-redundant inequalities implied by Theorem 2; (2) use 14 inequalities selected at random from the set of all 254 inequalities.

Figure 10 presents the sharp identified set and two examples of identified sets based on selecting random subsets of inequalities. The sharp identified set is relatively

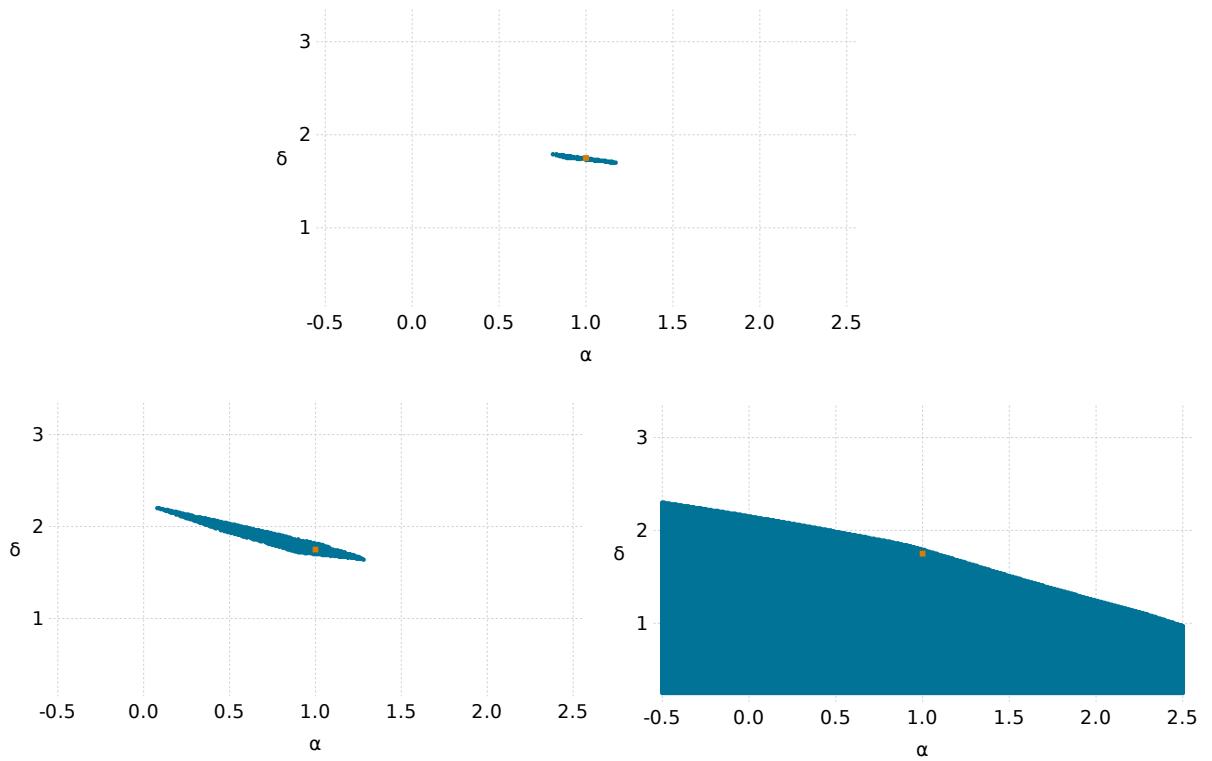


Figure 10: Simulated identified sets.

Comments: Upper panel depicts the sharp identified set. Lower left panel shows a “median” result of selecting inequalities at random. Lower panel shows a particularly bad combination of inequalities, selected at random.

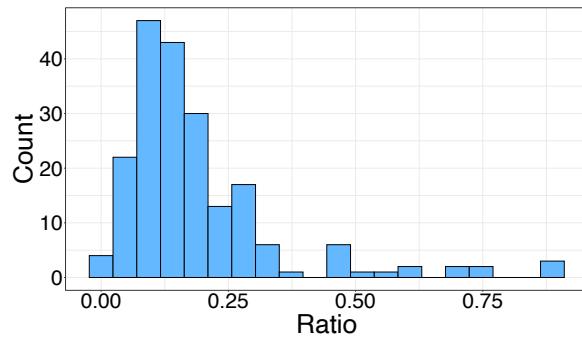


Figure 11: Size of the sharp identified set relative to the identified set constructed with the same number of inequalities selected at random.

tight, while identified sets based on other equally-sized collections of inequalities may be substantially larger. Figure 11 shows the distribution of the size (i.e., number of points on the grid that satisfy the inequalities) of the sharp identified set relative to the identified set produced with randomly selected inequalities. While it is infeasible to try all combinations of 14 out of 254 inequalities, the result can be interpreted as an empirical distribution of the quantity of interest under i.i.d. uniform sampling from a finite population. With overwhelming probability, the size of the sharp identified set is less than quarter of the size of the identified set constructed with a random subset of inequalities. This result confirms the suspicion that selecting inequalities based on experience with other models, or intuition, may result in a substantial loss of information.

6 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, I derived a simple characterization for constructing the smallest possible core-determining class in a general setting and illustrated its utility in several popular applications. The results can be applied broadly and 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.

References

- ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite dimensional analysis*. Springer.
- 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.
- 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 (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.
- BERESTEANU, A., AND F. MOLINARI (2008): “Asymptotic properties for a class of partially identified models,” *Econometrica*, 76(4), 763–814.
- 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.
- 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., 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.
- 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.
- 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.
- MATZKIN, R. L. (2007): “Nonparametric identification,” *Handbook of econometrics*, 6, 5307–5368.
- 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 non-negative 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): “Microeconomics with partial identification,” *Handbook of econometrics*, 7, 355–486.
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2015): “Moment inequalities and their application,” *Econometrica*, 83(1), 315–334.
- 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 Auxiliary Lemma

Lemma A.1. *Let $(\mathcal{U}, \mathcal{F}, P)$ be a probability space and $f_n : \mathcal{U} \rightarrow \mathbb{R}$ a sequence of bounded measurable functions with $|f_n| < M < \infty$ for all n . Suppose additionally that $\lim_{n \rightarrow \infty} \int f_n dP = 0$ and $\liminf_{n \rightarrow \infty} f_n \geq 0$, P -a.s. Then, there exists a subsequence n' such that $\lim_{n' \rightarrow \infty} f_{n'} = 0$, P -a.s.*

Proof. First, suppose that $\liminf_{n \rightarrow \infty} f_n = f$ with $P(f > 0) > 0$. Then, by Fatou's Lemma:

$$M = \liminf_{n \rightarrow \infty} \int (f_n + M) dP \geq \int \liminf_{n \rightarrow \infty} (f_n + M) dP > M,$$

which is a contradiction. Therefore, $\liminf_{n \rightarrow \infty} f_n = 0$, P -almost surely.

Next, write $f_n = f_n^+ - f_n^-$ where $f_n^+ = \max(f_n, 0)$ and $f_n^- = -\min(f_n, 0)$ defined pointwise and note that the sets on which f_n^+ and f_n^- are non-zero are non-overlapping. Therefore, $\liminf_{n \rightarrow \infty} f_n = 0$ implies that $\liminf_{n \rightarrow \infty} (-f_n^-) = 0$ or, equivalently, $\limsup_{n \rightarrow \infty} f_n^- = 0$. Since $\int f_n^+ dP$ and $\int f_n^- dP$ are bounded sequences and their sum converges to zero, one can choose a subsequence n' such that $\int f_{n'}^+ dP \rightarrow c$ and $\int f_{n'}^- dP \rightarrow c$. If $c > 0$, by the reverse Fatou's lemma,

$$0 < \limsup_{n' \rightarrow \infty} \int f_{n'}^- dP \leq \int \limsup_{n' \rightarrow \infty} f_{n'}^- dP = 0,$$

which is a contradiction. Therefore, $\int |f_{n'}| dP = \int (f_{n'}^+ + f_{n'}^-) dP \rightarrow 0$ so $f_{n'}$ converges to zero in $L_1(P)$. As is well known, there is a further subsequence convergent almost surely.

■

A.2 Proof of Theorem 1

Recall that \mathbf{U}_G denotes the class of all measurable unions of elements of the support of G , and \mathfrak{F} denotes the set of all closed subsets of \mathcal{Y} . By the arguments preceding the theorem, the class \mathbf{U}_G is core-determining, so that $\mathcal{M}(\mathbf{U}_G) = \mathcal{M}(\mathfrak{F})$. Now, suppose that for some collection of sets \mathcal{C} , $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathbf{U}_G)$. The first goal is to show that all sets $\tilde{A} \in \mathbf{U}_G \setminus \mathcal{C}$ must fail conditions (2) and/or (3) of the theorem.

Let $B(\mathcal{Y})$ denote the vector space of all bounded functions on \mathcal{Y} and consider

the space $(B(\mathcal{Y}) \times \mathbb{R}, \tau_{pw})$ endowed with the topology of pointwise convergence. It is a locally convex topological vector space. Consider a convex cone consisting of all finite linear combinations of vectors:

$$\left(\left\{ \begin{pmatrix} \mathbf{1}_A \\ -c_A \end{pmatrix} : A \in \mathcal{C} \right\}, \begin{pmatrix} \mathbf{1}_{\mathcal{Y}} \\ -1 \end{pmatrix}, \begin{pmatrix} -\mathbf{1}_{\mathcal{Y}} \\ 1 \end{pmatrix}, \left\{ \begin{pmatrix} \mathbf{1}_R \\ 0 \end{pmatrix} : R \in \mathcal{R} \right\} \right),$$

where \mathcal{R} is any distribution-determining class, and $c_A = C_G(A)$. A typical element of this cone can be expressed as:

$$\begin{pmatrix} f \\ c \end{pmatrix} = \sum_{l=1}^L \lambda_l \begin{pmatrix} \mathbf{1}_{A_l} \\ -c_l \end{pmatrix} + \lambda \begin{pmatrix} \mathbf{1}_{\mathcal{Y}} \\ -1 \end{pmatrix} + \sum_{k=1}^K \gamma_k \begin{pmatrix} \mathbf{1}_{R_k} \\ 0 \end{pmatrix}$$

for some $\{\lambda_l\} \geq 0$, $\lambda \in \mathbb{R}$, $\{\gamma_k\} \geq 0$, and $L, K \in \mathbb{N}$.

Pick any $\tilde{A} \in \mathsf{U}_G \setminus \mathcal{C}$ and consider a vector $(\mathbf{1}_{\tilde{A}}, -c_{\tilde{A}})$. By Farkas Lemma (see Corollary 5.84 in [Aliprantis and Border, 2006](#)), either this vector belongs to the pointwise closure of the above cone, or there is a continuous linear operator $L : B(\mathcal{Y}) \times \mathbb{R} \rightarrow \mathbb{R}$ that separates the vector from the cone in a sense that $L(f, c) \geq 0$ for any vector (f, c) in the cone and $L(\mathbf{1}_{\tilde{A}}, -c_{\tilde{A}}) < 0$. If the separating operator exists, it takes the form $L(f, c) = \nu(f) + \gamma c$, where $\nu : (B(\mathcal{Y}), \tau_{pw}) \rightarrow \mathbb{R}$ is a continuous linear operator and $\gamma \in \mathbb{R}$, and satisfies $\nu(\mathbf{1}_{\tilde{A}}) < \gamma c_{\tilde{A}}$, $\nu(\mathbf{1}_A) \geq \gamma c_A$, for all $A \in \mathcal{C}$, $\nu(\mathbf{1}_{\mathcal{Y}}) = \gamma$, and $\nu(\mathbf{1}_R) \geq 0$, $\forall R \in \mathcal{R}$. Note that $\gamma > 0$ or else $\nu = 0$ and separation would be impossible. Then, it is easy to verify that the set function $\mu(A) \equiv \gamma^{-1}\nu(A)$ defines a probability measure on \mathcal{Y} that satisfies $\mu(\tilde{A}) < c_{\tilde{A}}$ and $\mu(A) \geq c_A$ for all $A \in \mathcal{C}$. This contradicts the assumption $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathsf{U}_G)$, so the separating operator cannot exist. Therefore, the vector $(\mathbf{1}_{\tilde{A}}, -c_{\tilde{A}})$ must belong to the pointwise closure of the cone defined above.

In the topology of pointwise convergence, closures of sets must generally be described with nets rather than sequences (see, e.g., Section 2.15 in [Aliprantis and Border, 2006](#)). For the purposes of this proof, however, the distinction between nets and sequences is purely notational, so I proceed with sequences. In view of the above, for some sets $A_{l,N} \in \mathcal{C}$ and $R_{k,N} \in \mathcal{R}$ and constants $\lambda_{l,N} \geq 0$, $\lambda_N \in \mathbb{R}$ and $\gamma_{k,N} \geq 0$,

$$\mathbf{1}(y \in \tilde{A}) = \lim_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(y \in A_{l,N}) + \lambda_N + \sum_{k=1}^{K_N} \gamma_{k,N} \mathbf{1}(y \in R_{k,N}) \right) \quad (\text{A.1})$$

for all $y \in \mathcal{Y}$, and

$$P(G \subseteq \tilde{A}) = \lim_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} P(G \subseteq A_{l,N}) + \lambda_N \right). \quad (\text{A.2})$$

Note that it must be either $\lambda_N \rightarrow 0$, or there is a $\kappa > 0$ such that $\lambda_N < -\kappa$ for large enough N . Indeed, assuming that $\lambda_N > \kappa > 0$ for all N contradicts (A.1) for any $y \notin \tilde{A}$. Without loss of generality, one can choose a sequence such that all $\lambda_{l,N} > \xi > 0$, and $\max(\lambda_{l,N}, |\lambda_N|, \gamma_{k,N}) < M < +\infty$ for large enough N . From here, the proof proceeds in three steps.

Step 1 First, I show that along a subsequence, still denoted by N for simplicity,

$$\mathbf{1}(G(u) \subseteq \tilde{A}) = \lim_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(G(u) \subseteq A_{l,N}) + \lambda_N \right), \quad (\text{A.3})$$

for all $u \in \tilde{\mathcal{U}} \subseteq \mathcal{U}$ with $P(\tilde{\mathcal{U}}) = 1$, and, for all $y \in \tilde{\mathcal{Y}} = \bigcup\{G(u) : u \in \tilde{\mathcal{U}}\}$,

$$\mathbf{1}(y \in \tilde{A}) = \lim_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(y \in A_{l,N}) + \lambda_N \right). \quad (\text{A.4})$$

Note that for any fixed $u \in \mathcal{U}$,

$$\begin{aligned} \mathbf{1}(G(u) \subseteq \tilde{A}) &= \min_{y \in G(u)} \mathbf{1}(y \in \tilde{A}) = \min_{y \in G(u)} \limsup_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(y \in A_{l,N}) + \lambda_N \right) \\ &\geq \limsup_{N \rightarrow \infty} \min_{y \in G(u)} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(y \in A_{l,N}) + \lambda_N \right) \\ &\geq \limsup_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \min_{y \in G(u)} \mathbf{1}(y \in A_{l,N}) + \lambda_N \right) = \limsup_{N \rightarrow \infty} \left(\sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(G(u) \subseteq A_{l,N}) + \lambda_N \right), \end{aligned}$$

and therefore

$$\liminf_{N \rightarrow \infty} \left(\mathbf{1}(G(u) \subseteq \tilde{A}) - \sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(G(u) \subseteq A_{l,N}) - \lambda_N \right) \geq 0.$$

Moreover, Equation (A.2) can be written as:

$$\lim_{N \rightarrow \infty} \int \left(\mathbf{1}(G(u) \subseteq \tilde{A}) - \sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(G(u) \subseteq A_{l,N}) - \lambda_N \right) dP(u) = 0.$$

By Lemma A.1, the integrand in the above display converges to zero almost surely along a subsequence, which shows (A.3). Now, suppose that (A.4) holds strictly for some $y \in G(u)$ for $u \in \tilde{\mathcal{U}}$. If $G(u) \subseteq \tilde{A}$,

$$\begin{aligned} \mathbf{1}(G(u) \subseteq \tilde{A}) &= \limsup_{N \rightarrow \infty} \sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(G(u) \subseteq A_{l,N}) + \lambda_N \\ &\leq \limsup_{N \rightarrow \infty} \sum_{l=1}^{L_N} \lambda_{l,N} \mathbf{1}(y \in A_{l,N}) + \lambda_N \\ &< \mathbf{1}(y \in \tilde{A}), \end{aligned}$$

which states that $1 < 1$, a contradiction. If $G(u) \not\subseteq \tilde{A}$, the same argument applied to any $y \in G(u) \cap \tilde{A}^c$ leads to $0 < 0$, a contradiction. Therefore, (A.4) follows.

Step 2 Next, I show that for all $u \in \tilde{\mathcal{U}}$,

- (i) If $G(u) \subseteq \tilde{A}$, then, for all N large enough, for each $l \leq L_N$, either $G(u) \subseteq A_{l,N}$ or $G(u) \cap A_{l,N} = \emptyset$;
- (ii) If $G(u) \cap \tilde{A}^c \neq \emptyset$, then, for all N large enough, for each $l \leq L_N$, either $(G(u) \cap \tilde{A}^c) \subseteq A_{l,N}^c$ or $(G(u) \cap \tilde{A}^c) \cap A_{l,N}^c = \emptyset$.

Consider any $u \in \tilde{\mathcal{U}}$ such that $G(u) \subseteq \tilde{A}$ and $y \in G(u)$, and subtract (A.3) from (A.4).

$$\lim_{N \rightarrow \infty} \sum_{l=1}^{L_N} \lambda_{l,N} (\mathbf{1}(y \in A_{l,N}) - \mathbf{1}(G(u) \subseteq A_{l,N})) = 0,$$

Since $\lambda_{l,N} > 0$ and the difference of indicators is non-negative, it must be that $\mathbf{1}(y \in A_{l,N}) = \mathbf{1}(G(u) \subseteq A_{l,N})$ for all large enough N . This implies (i). Now, consider any $u \in \tilde{\mathcal{U}}$ such that $G(u) \cap \tilde{A}^c \neq \emptyset$ and any $y \in G(u) \cap \tilde{A}^c$. Adding and subtracting one inside the parentheses, the above display can be written as:

$$\lim_{N \rightarrow \infty} \sum_{l=1}^{L_N} \lambda_{l,N} (\mathbf{1}(y \in A_{l,N}^c) - \mathbf{1}(G(u) \cap A_{l,N}^c \neq \emptyset)) = 0.$$

Since $\lambda_{l,N} > 0$, the above implies that $\mathbf{1}(y \in A_{l,N}^c) = \mathbf{1}(G(u) \cap A_{l,N}^c \neq \emptyset)$ for all large enough N . That is, $G(u) \cap A_{l,N}^c \neq \emptyset$ if and only if $y \in A_{l,N}^c$ for all $y \in G(u) \cap \tilde{A}^c$. This implies (ii).

Step 3 Finally, I demonstrate that the set \tilde{A} fails to satisfy either condition (2) or condition (3) in the statement of the theorem. It suffices to consider two cases.

First, suppose that $\lambda_N = 0$ for all N along a subsequence, or $\lambda_N \rightarrow 0$ as $N \rightarrow \infty$. Then, Equation (A.4) and the fact that all $\lambda_{l,N}$ are bounded away from zero imply that $\tilde{A} \cap \tilde{\mathcal{Y}} = \lim_{N \rightarrow \infty} \bigcup_{l=1}^{L_N} A_{l,N} \cap \tilde{\mathcal{Y}}$. By (i) above, for all $u \in \tilde{\mathcal{U}}$ such that $G(u) \subseteq \tilde{A}$, either $G(u) \subseteq A_{l,N}$ or $G(u) \subseteq \tilde{A} \setminus A_{l,N}$ must hold, for each $l \leq L_N$ eventually. This implies that all $A_{l,N}$ are disjoint and there is no $u \in G^-(\tilde{A}) \cap \tilde{\mathcal{U}}$ such that $G(u)$ connects $A_{l,N}$ and $\tilde{A} \setminus A_{l,N}$. Thus, condition (2) of the theorem fails.

Second, suppose that $\lambda_N < -\kappa$ for all N along a subsequence for some $\kappa > 0$. Then, Equation (A.4) and the fact that $\lambda_{l,N}$ are bounded away from zero imply that $\tilde{\mathcal{Y}} = \lim_{N \rightarrow \infty} \bigcup_{l=1}^{L_N} A_{l,N} \cap \tilde{\mathcal{Y}}$. There are three options: (1) $\tilde{A} \cap A_{l^*,N} \neq \emptyset$ and $\tilde{A} \cap A_{l^*,N}^c \neq \emptyset$ for some $l^* = l^*(N)$ infinitely often; (2) for some l , $\tilde{A} \subseteq A_{l,N}$ eventually, and there exist $l^* = l^*(N)$ such that $\tilde{A} \cap A_{l^*,N} = \emptyset$ infinitely often; and (3) $\tilde{A} \subseteq A_{l,N}$ for all $l \leq L_N$ eventually.

Consider option (1) and pick any $u \in G^-(\tilde{A}) \cap \tilde{\mathcal{U}}$. By (i) above, either $G(u) \subseteq A_{l,N}$ or $G(u) \cap A_{l,N} = \emptyset$ must hold for each $l \leq L_N$ eventually. If $\tilde{A} \setminus A_{l,N} \neq \emptyset$ for some l, N , there is no $u \in G^-(\tilde{A}) \cap \tilde{\mathcal{U}}$ such that $G(u)$ connects $A_{l,N}$ and $\tilde{A} \setminus A_{l,N}$. Then, condition (2) of the theorem fails. Now, consider option (2) and pick any $u \in \tilde{\mathcal{U}}$ such that $G(u) \subseteq \tilde{A}^c$. By observation (ii), $G(u) \subseteq A_{l,N}$ or $G(u) \subseteq A_{l,N}^c$ for each $l \leq L_N$ eventually. This implies that the subgraph induced by $(A_{l^*,N}, G^-(A_{l^*,N}))$ is disconnected from the rest of the graph, P -a.s., which contradicts the assumption of the theorem. It remains to consider option (3). Note that (A.4) implies that $\tilde{A}^c \cap \tilde{\mathcal{Y}} = \lim_{N \rightarrow \infty} \bigcup_{l=1}^{L_N} A_{l,N}^c \cap \tilde{\mathcal{Y}}$. Pick any $u \in \tilde{\mathcal{U}}$ such that $G(u) \cap \tilde{A}^c \neq \emptyset$. By (ii) above, either $(G(u) \cap \tilde{A}^c) \subseteq A_{l,N}^c$ or $(G(u) \cap \tilde{A}^c) \cap A_{l,N}^c = \emptyset$ must hold for each $l \leq L_N$ eventually. In other words, there is no $u \in G^{-1}(\tilde{A}^c) \cap \tilde{\mathcal{U}}$ such that $G(u)$ connects A_l^c and $A_j^c \setminus A_l^c$. Then, condition (3) of the theorem fails.

Thus, if \mathcal{C} is any core-determining class and $A \in \mathbb{U}_G \setminus \mathcal{C}$, then A must fail to satisfy condition (2) or (3) of the theorem with $A_1, A_2 \in \mathcal{C}$. Conversely, if A satisfies conditions (2) and (3), then any class \mathcal{C} that does not include A cannot be core-determining. This means that A is non-redundant. Additionally, note that the above

argument implies that a set $A \in \mathbf{U}_G$ is redundant given sets $A_1, A_2 \in \mathbf{U}_G$ if and only if it fails to satisfy conditions (2) or (3) of the theorem for these A_1, A_2 .

Finally, I show that the class of all non-redundant sets is core-determining. Consider removing all redundant sets from \mathbf{U}_G step by step. First, remove any set A that can be expressed as $A = \bigcup_{n \geq 1} A_n$ with $A_i \cap A_j = \emptyset$, and $G^-(A_i) \cap G^-(A_j) = \emptyset$ for all $i \neq j$. Then, remove any set A such that $A^c = \bigcup_{n \geq 1} A_n^c$ where $A_i^c \cap A_j^c = \emptyset$ and $G^{-1}(A_i^c) \cap G^{-1}(A_j^c) = \emptyset$. The above procedure may lead to logical inconsistencies if one must use A to show the redundancy of A' but also use A' to show the redundancy of A . I will show that this is impossible. Indeed, suppose that A fails condition (2) of the theorem so that $A = A' \cup A''$ with $A' \cap A'' = \emptyset$, and $G^-(A) = G^-(A') \cup G^-(A'')$, but also $A' = A \cap A'''$ with $A \cup A''' = \mathcal{Y}$ and $G^-(A) \cup G^-(A''') = \mathcal{U}$. Then, $(A''')^c = A''$, and there are no u connecting A'' with either A' or A^c , meaning that the subgraph induced by $(A'', G^-(A''))$ is disconnected from the rest of the graph, P -a.s. This contradicts the assumption of the theorem. The case when A fails condition (3) of the theorem can be considered similarly. The proof is now complete. ■

A.3 Proof of Theorem 2

The result follows directly from Theorem 1 and the observations preceding the statement of the theorem.

A.4 Size of the Smallest Core-Determining Class

Let $\mathcal{U} = \{u_1, \dots, u_K\}$, $\mathcal{Y} = \{y_1, \dots, y_S\}$, 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 $y_1 \preccurlyeq_{\mathcal{Y}} \dots \preccurlyeq_{\mathcal{Y}} y_S$. 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 $\{y_m, y_{m+1}, \dots, y_{m+l}\}$, 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 A.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 $\mathcal{A}_{\emptyset} = \{S \in \mathcal{S} : G^-(S) = \emptyset\}$, 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 $A_l = (\overline{S}_l, \underline{S}_{l+1}) \neq \emptyset$.
2. Each interior S_l and A_l is in $\mathcal{U}_{\mathcal{A}_{NM} \cup \mathcal{A}_{\emptyset}}$.
3. If $L \geq 2$, $G^-(S) \cap \mathcal{U}_{NC} \neq \emptyset$.

Then, \mathcal{C} is core-determining. The smallest core-determining class can be computed by applying Theorem 2 to \mathcal{C} .

Despite a lengthy statement, the idea of Theorem A.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 A.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 A.1.1. *Let \mathcal{U} and \mathcal{Y} be finite sets and $G : \mathcal{U} \Rightarrow \mathcal{Y}$ 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 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, as discussed below.

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 $m = 1, \dots, M$, apply the greedy algorithm:
 - (1) Let $y_m^1 = y_m$ and for each $l \geq 1$, 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 $\mathcal{Y}_m^{1:l} = \{y_m^1, \dots, y_m^l\}$. 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 $\succ_{\mathcal{Y}}^m$.
 - (3) Reorder the elements of \mathcal{U} so that $u \succ_{\mathcal{U}}^m u'$ if $\min G(u) \succ_{\mathcal{Y}}^m \min G(u')$ and $\max G(u) \succ_{\mathcal{Y}}^m \max G(u')$ and let \mathbf{B}_G^m denote the resulting bipartite graph.
3. Among all $\{\mathbf{B}_G^m : m = 1, \dots, M\}$, select the “most monotone one” in the sense defined below.

Let \mathbf{B} denote a bipartite graph with vertices in $\mathcal{U} = \{u_1, \dots, u_K\}$ and $\mathcal{Y} = \{y_1, \dots, y_S\}$ 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 $\mathbf{E} \subseteq \{1, \dots, K\} \times \{1, \dots, S\}$. 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 \succ_C on $\mathcal{I}(\mathbf{B})$ as:

$$\mathbf{B}' \succ_C \mathbf{B} \iff \begin{cases} \min_s \mathbf{E}'_k, \geq \min_s \mathbf{E}_k & \text{for all } k = 1, \dots, K \\ \max_s \mathbf{E}'_k \leq \max_s \mathbf{E}_k \\ \min_k \mathbf{E}'_s \geq \min_k \mathbf{E}_s & \text{for all } s = 1, \dots, S \\ \max_k \mathbf{E}'_s \leq \max_k \mathbf{E}_s \end{cases}$$

where the inequalities are interpreted component-wise.

A.4.1 Proof of Theorem A.1

Below, I refer to segments contiguous with respect to $\succ_{\mathcal{Y}}$ 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, I will show that $\mathcal{C}^* \subseteq \mathcal{C}$, where \mathcal{C}^* is the smallest core-determining class characterized in Corollary 2. 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 $S \in \mathcal{U}_G$. Combining the above observations yields $S \in \mathcal{U}_{\mathcal{A}_{NM}}$. 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 $A_l = (\overline{S}_l, \underline{S}_{l+1}) \neq \emptyset$. 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 $G^-(S_l) = \emptyset$. The same applies to each A_l , with the addition that connectedness of $(S, G^-(S))$ requires $G^-(S) \cap \mathcal{U}_{NS} \neq \emptyset$. Therefore, the class \mathcal{C} defined in the statement of the theorem is core-determining.