# Identification of Incomplete Preferences

Arie Beresteanu* Luca Rigotti†

August, 2021‡

**Abstract**

We provide a sharp identification region for discrete choice models in which consumers' preferences are not necessarily complete and only aggregate choice data is available to the analysts. Behavior with non complete preferences is modeled using an upper and a lower utility for each alternative so that non-comparability can arise. The identification region places intuitive bounds on the probability distribution of upper and lower utilities. We show that the existence of an instrumental variable can be used to reject the hypothesis that all consumers' preferences are complete, while attention sets can be used to rule out the hypothesis that all individuals cannot compare any two alternatives. We apply our methods to data from the 2018 mid-term elections in Ohio.

**Keywords**: Partial Identification, Incomplete Preferences, Vagueness, Knightian Uncertainty, Random Sets.

*Department of Economics, University of Pittsburgh, arie@pitt.edu.

†Department of Economics, University of Pittsburgh, luca@pitt.edu.

‡We thank Miriam Blumenthal for excellent RA work, Pawel Dziewulski for very careful comments on an early draft, Ariel Rubinstein and Richard Van Weelden, as well as audiences at Cornell, Duke, Penn State, Pitt, SUNY Albany, Tel Aviv, and the RUD conference.

# 1 Introduction

Since McFadden's 1974 paper, discrete choice models have been one of the cornerstones of applied economics analysis. These models, like much of economics, assume that individuals' preferences are complete: individuals can always rank any pair of alternatives. Doing so in identification problems is convenient because completeness induces a ranking between any pair of alternatives, and therefore makes identification of rational behavior easier.[1] In this paper we tackle identification when individuals' preferences are allowed to be incomplete, individuals make only one choice, and only aggregate behavior is observable. This is a classic discrete choice setting modified so that completeness does not necessarily hold. The objective is to identify properties of the probability distribution of preferences across heterogeneous individuals when aggregated choice data for the population is observed.

When preferences are not complete there are pairs of alternatives individuals may not be able to rank. In this setting, observers who see a particular alternative being chosen cannot infer that this alternative must have been at least as good as those not chosen. They can only infer that the alternatives not chosen could not have been strictly preferred to the chosen one. When using data to learn about the distribution of preferences in the population, allowing for incompleteness poses what may seem like an insurmountable problem. Because the theory says nothing about how choices between incomparable alternatives are made, data cannot help distinguish between choices made because of a definite preference from choices made randomly because two alternatives are not comparable. In standard discrete choice models this source of randomness is ruled out by assuming everyone's preferences are complete, and one may suspect that when completeness is not imposed data cannot say anything about preferences. To the contrary, we show that also in this case data can be used to learn about the distribution of preference parameters. Obviously, identification has weaker properties than in the case of complete preferences.

Our first main result shows that although point identification is not possible, partial identification is. In particular, we characterize the sharp identification region when two or more options are in the alternatives set. For example, in binary choice we show that for each alternative the probability that it is preferred over the other alternative must always be smaller than the fraction of consumers who choose that alternative. Intuitively, point identification is not possible because individuals who make a particular choice may do so when that choice is not comparable to its alternative. On the other hand, if a given fraction of consumers chose a certain alternative, it cannot be that the probability that this alternative has higher utility exceeds the observed

[1] Indifference implies alternatives are ranked as equal, and is typically deemed a zero probability occurrence.

fraction. We also show how this logic extends to settings where more than two alternatives are available.

The identification region we characterize does not rule out the possibility that all individuals have complete preferences. This is natural because allowing for incompleteness is not equivalent to ruling out completeness. Our second main result gives intuitive sufficient conditions for an observer to conclude that a population must include at least some individuals whose preferences are not complete. Suppose one can find a random variable that influences observed aggregate choices but is orthogonal to individual preferences. If such an 'instrumental variable' exists, one can conclude that at least some individuals' preferences cannot be complete. In our application section, we show how the order in which candidates are presented on an election ballot can be used as an instrument to infer that some voters' preferences cannot be complete.

We consider a setting in which data is collected about the choices of a population of heterogeneous agents who must choose one of a given number of alternatives. In the standard model with complete preferences, each alternative is associated with a utility value that is known to the decision maker but not known to the analyst. The analyst, on the other hand, knows from data the frequency with which each alternative is chosen. Observed choices are used to make inferences about the probability with which one option is preferred to the others in the population. Inference is driven by the assumption that choices are made in a rational, utility maximizing, fashion. Observing the proportion of individuals who chose one alternative implies that same proportion must assign higher utility to the corresponding alternative. We model incompleteness by allowing for the possibility that each alternative is associated with many utility values. In this setting, alternatives are ranked using the two most extreme of these values: upper and lower utility. An alternative is preferred if its lower utility is higher than the upper utility of the other. Rationality, however, implies an alternative can be chosen as long as its higher utility is higher than the lower utility of the other. A simple example of this framework is given by the *interval order* of Fishburn (1970), where each alternative is associated with an interval of utilities.[2] When two intervals are disjoint, the corresponding alternatives can be ranked and rationality determines choice; when two intervals overlap, the corresponding alternatives are not comparable and choice is indeterminate. Therefore, the link between data about observed choices and the probability that alternatives are preferred is weakened once incompleteness is allowed.

Using methods from random set theory and results from the literature on partial identifica-

[2]Examples when uncertainty or ambiguity are present are given by Aumann (1962) and Dubra et al. (2004) in a multi utility framework, Bewley (2002) in a multi probability environment, and more recently Echenique et al. (2021) that combines Fishburn's interval utility model with Bewley's, and Miyashita and Nakamura (2021).

tion, we connect observed data and the probability that an alternative is preferred by looking at the probability distribution of upper and lower utilities in the population. The proportion of individuals who choose an alternative implies a bound on the probability that an individual prefers that alternative. Intuitively, the bounds follow from two observations. First, the fraction of individuals who chooses an alternative gives the upper bound of the probability that a randomly chosen individual assigns a higher interval to it. Second, the data cannot impose a lower bound to this same probability as it could be that all those who make some choice do it because they cannot compare it to the other option. These two bounds yields the sharp identification region.

Without additional assumptions, the identification region does not rule out the possibility that all individuals who choose an alternative do so because it is ranked better, and thus non-comparability never occurs. To rule out this possibility, we assume that an additional random variable is observed. This random variable has the property of being independent of preferences and yet influencing choice when alternatives cannot be compared. Intuitively, if the fraction of individuals who choose a certain alternative changes with the realization of this random variable, it must be that some individuals were choosing randomly because their preferences are not complete. This random variable thus rules out the possibility that all individuals have complete preferences because its existence makes the identification region smaller.

We study several extensions of the basic model that each can make the identification region smaller. First, we examine a model in which a known fraction of individuals pays attention only to a subset of the alternatives. Second, we show how minimizing ex-post regret could bring point identification. Third, an extension that is particularly relevant for our application takes care of the situation in which the choices of a known fraction of the population are not observed; for example, a known fraction of individuals go to the polls yet abstain in some of the races on the ballot. Finally, while our results are cast in a world without randomness, we show that one can allow for ambiguity by extending our framework to study identification of the preferences described in the Knightian Decision Theory of Bewley (1986).[3] This extension shows that the crucial part of our analysis is not the interval order model but rather the notion that rational behavior is driven by two separate thresholds. In our empirical application we demonstrate how several behavioral assumptions can be combined to reduce the size of the identification set.

We apply our methods to precinct-level voting data from the 2018 midterm elections in Ohio, focusing on the two races for the Ohio Supreme Court in Lorain county. This empirical

[3] Although we do not pursue it, a similar exercise could be performed in the multiple utility framework of Aumann (1962).

application fits the theory in several ways. First, individual choices are not observable like in our setting. Second, the phenomenon of roll-off voting can be thought of as an example through which incompleteness can directly manifest itself. In the two races we consider party affiliation is not on the ballot (even though candidates were selected by each party in their primary elections), and more than 20% of the turnout voters did not vote in these races even though they went to the polls. Third, election data helps us illustrate the role of attention sets by assuming that partisan voters only consider candidates of their own party. We implement this using voter registration files for each precinct. Finally, Ohio election rules provide a good example for an instrumental variable: the order in which the candidates are presented on the ballot changes from precinct to precinct. Candidates' order on the ballot is unlikely to be correlated with voters' preferences, while it can have an effect on vote shares, particularly when each candidate's party affiliation is not on the ballot as is the case for Judicial elections in Ohio. We show that the order has a significant, but small, effect on the identification region in our races.

Our paper brings together two strands of recent literature: the one focused on the consequences of incomplete preferences for economic outcomes, and the one focused on the econometric identification of economic models that have non-unique predictions. The former dates back to Luce (1956), who speaks about intransitive indifference and Aumann (1962) who introduce incompleteness and multiple utility functions for a decision maker with objective probabilities (see Bayrak and Hey (2020) for a recent survey of some of this literature). To the best of our knowledge, we are the first to tackle econometric identification of these models from aggregate data, and we are the first to connect preference incompleteness with partial identification.[4]

Our approach to the econometric identification of the probability distribution of incomplete preferences from aggregate data is inspired by a recent trend toward weaker assumptions in econometric modeling. Manski (2003) coins the term *The Law of Decreasing Credibility* saying that "*The credibility of inference decreases with the strength of the assumptions maintained*". Seeking more credible results, a researcher can either weaken assumptions on the behavior of decision makers in the model or weaken assumptions on the data generating process. We focus on the first aspect by dropping completeness, while also trying to make few assumptions about the data. Econometricians and practitioners have long recognized that using Probit and Logit to estimate preferences involves maintaining a very restrictive set of both behavioral and econometric assumptions. The attempts to generalize these estimators focused mostly on relaxing the latter. A range of non-parametric and semi-parametric models of discrete choice have been proposed

[4]Dziewulski (2021) uses revealed preference arguments to recover an individual's preference ordering with intransitive indifferences.

(see Matzkin (1994) and Matzkin (2007) for a discussion). In particular, partial identification in discrete choice may occur when some variables are not precisely measured. Manski and Tamer (2002) and Beresteanu et al. (2012) consider a discrete choice model where some characteristics of the alternatives are interval valued. Horowitz and Manski (1998) consider a situation where some choices and some characteristics are unobserved due to survey non response. Horowitz and Manski (1995) consider the case where data are corrupted or contaminated. We differ from these papers because partial identification in our case stems from a behavioral assumption of lack of completeness in the preferences.

In spirit, this paper is related to behavioral models where rationality is weakened usually focus on finding tests for rationality in the general sense as in Kitamura and Stoye (2018) and by Hoderlein (2011). These papers look for violations of utility maximizing behavior in data on individual choices. Violations, if found, are not attributed to failure of any specific assumption but rather to a general lack of rationality by some individuals. In our case, decision makers are perfectly rational even though their preferences are not complete. Finally, our results on attention sets are related to theories introduced in Masatlioglu et al. (2012) and Eliaz and Spiegler (2011). Manzini and Mariotti (2014)) introduce assumptions on the way individuals restrict their attention to a subset of the alternatives such that the parameters of the model are point identified. To large extent the models of attention sets constrain the behavior of the individuals rather than relax the assumptions of the model. In addition to partial identification resulting from incomplete preferences, researchers can face non response as well. We show that the identification region can be written as a convex combination of an identification region resulting from incomplete preferences and an identification region resulting from non response.

The remainder of the paper is organized as follows. Section 2 introduces incomplete preferences and describes rational behavior. Section 3 discusses random decision makers and non-parametric identification of preferences distribution. Section 4 focuses on binary choice. Section 5 illustrates several extension of our basic framework. In section 6 we apply our findings to a data set on voting in Ohio, and show how this data can be used to illustrate all our findings. Section 7 concludes.

## 2 Incomplete Preferences and Rationality

Our objective is to analyze the standard discrete choice model without imposing the assumption that preferences are complete. In the standard model, completeness is reflected by the idea that choice between alternatives is driven by the utility assigned to each. This follows from the well known result that, with a finite number of alternatives, any complete and transitive preference relation has a utility function representing it. The utility function assigns a single number to each alternative (its utility), and the ordering of these numbers is used to rank alternatives and thus to model choices. Without completeness this is no longer the case: one cannot assign a unique utility value to each alternative. In this paper, we focus on the situation in which each alternative is associated with two numbers, and these two numbers are used to determine the ranking between alternatives, if such a ranking exists.

Let $X$ be the set of alternatives over which an individual's preference order $\succ$ is defined. For each $x \in X$ we assume there exist two real numbers $\bar{u}(x)$ and $\underline{u}(x)$, with $\bar{u}(x) \geq \underline{u}(x)$. We refer to these two numbers as an alternative's upper and lower utility respectively, and we use the word vagueness to describe the length of the utility interval $[\underline{u}(x), \bar{u}(x)]$ associated to each $x$. Upper and lower utility can be used to describe the individual's preference ordering between any two alternatives, and this preference can then be used to describe her behavior when faced with a particular subset of feasible alternatives. First, for any $x, y \in X$ we say that $x$ is (strictly) preferred to $y$ if and only if $\bar{u}(y) < \underline{u}(x)$. In words, one alternative is preferred if its lower utility exceeds the upper utility of the other. Two alternatives are not comparable when neither this inequality nor its opposite are satisfied (their utility intervals overlap). We say that two alternatives are indifferent if their upper and lower utilities are the same (their utility intervals are the same).

Next, we describe how choices are made from a given set of feasible alternatives $\mathcal{A} \subseteq X$ given the preference relation above. The main idea is that choice must allow for the possibility that some alternatives are not ranked. In particular, an alternative can be chosen from a set provided there is no other option in that set that is strictly preferred to it. Suppose only two alternatives are feasible, then one of them can be chosen if it is either preferred to the other, or if it is not comparable to it. In other words, an alternative can be chosen as long as it is not dominated. In general, when there are more than two feasible possibilities in $\mathcal{A}$, an alternative can be chosen provided there is no other option in $\mathcal{A}$ that is strictly preferred to it. We formalize these ideas using the following definition.

**Definition 1** *Given $\mathcal{A} \subseteq X$, the subset of* ***non-dominated alternatives*** *is defined as*

$$M(\mathcal{A}) = \{a \in \mathcal{A} : \forall b \in \mathcal{A},\ \bar{u}(a) \geq \underline{u}(b)\} \subseteq \mathcal{A}.$$

$M(\mathcal{A})$ is the set of all alternatives in $\mathcal{A}$ that are not (strictly) dominated by any element of $\mathcal{A}$. When preferences are complete, the set $M(\mathcal{A})$ includes only alternatives that are indifferent to each other. Without completeness, the set $M(\mathcal{A})$ may include several incomparable alternatives. The set of non-dominated alternatives is used to describe observable behavior. Let $y \in \mathcal{A}$ denote an observable choice (selection) from some feasible set $\mathcal{A}$; we assume the following.

**Assumption 1 (Rationality)** *$y \in M(\mathcal{A})$ for any $\mathcal{A}$.*

Rationality has implications for choice data: any observable choice must be an element of the set of non-dominated alternatives. Thus, the analyst knows that dominated alternatives cannot be chosen. When $M(\mathcal{A})$ is not a singleton rationality does not make unique predictions about behavior. When preferences are complete, typical assumptions imply that indifference is a zero probability event, and thus the analyst can treat $M(\mathcal{A})$ as a singleton. Without completeness, assuming that indifference is a zero probability event does not necessarily make $M(\mathcal{A})$ a singleton because incompleteness is not as knife-edge as indifference. One can rule out indifference between $x$ and $y$ by asking that the intervals $[\underline{u}(x), \bar{u}(x)]$ and $[\underline{u}(y), \bar{u}(y)]$ are different, but this is not enough to rule out any overlap between them. Rationality also reflects the idea that no selection rule is a-priory imposed among incomparable alternative because it only says that an alternative that is strictly dominated is never chosen.

Although we cast the paper in the language of upper and lower utilities for ease of exposition, our identification results apply to any incomplete preference relation that yields a set of undominated alternatives constructed using only two numbers as stated in Definition 1. There are many examples of such preferences, and we end this section by describing one of them as an illustration of how upper and lower utilities can obtain; in Section 5.2 we show that our approach is also consistent with the Knightian decision theory of Bewley (1986) where preferences are not complete because of ambiguity.[5]

## 2.1 Interval Orders and Vagueness

Interval orders are presented in Fishburn (1970) and are well suited to illustrate our setup. They can be obtained under simple assumptions on preferences. Let $(X, \succ)$ be a partially ordered set such that $X$ is a finite set and $\succ$ is a strict preference relation over $X$. The first property is irreflexivity: $\neg(x \prec x)$ (where $\neg$ denotes logical negation). The second property is a special form of transitivity: $x \prec y$ and $z \prec w \Rightarrow x \prec w$ or $z \prec y$.[6] Fishburn calls a preference that

[5] Other examples could be Aumann (1962) and Dubra et al. (2004), or the more recent Echenique et al. (2021) and Miyashita and Nakamura (2021).

[6] One can easily verify that these two properties imply the usual transitivity.

satisfies these two properties an **interval order**, and he shows that $\succ$ can be represented using two functions as the following theorem illustrates.

**Theorem** [Fishburn (1970)] If $\succ$ is an interval order and $X$ is finite, then there exists two functions $\underline{u} : X \to \mathbb{R}$ and $\sigma : X \to \mathbb{R}$ with $\sigma(x) > 0$ for all $x \in X$, such that

$$x \prec y \qquad \text{if and only if} \qquad \underline{u}(x) + \sigma(x) < \underline{u}(y).$$

Thus, $y$ is preferred to $x$ if and only if the utility of $y$ exceeds the utility of $x$ by some strictly positive amount. One can think of each alternative as being associated with an interval on the real line. The lower bound of that interval represents its utility, while the width of that interval represents imprecision in that utility. In this spirit, Fishburn calls $\sigma$ the *vagueness* function since it measures the amount of imprecision in the utility associated with each alternative. The name interval order follows from the observation that alternatives are compared using intervals: when $y$ is preferred to $x$ the interval $[\underline{u}(y), \underline{u}(y) + \sigma(y)]$ lies to the right of the interval $[\underline{u}(x), \underline{u}(x) + \sigma(x)]$. Interval orders are clearly not necessarily complete. When the two intervals overlap, $x$ and $y$ are not comparable. In this setting, although strict preference is transitive, non-comparability is not. In other words, $x$ may be not comparable to $y$, and $y$ maybe not comparable to $z$, but $z$ is strictly preferred to $x$. This idea is sometimes referred to as intransitive indifference.[7]

Interval orders easily map to our framework by letting $\bar{u}(x) = \underline{u}(x) + \sigma(x)$ so that each alternative is associated with a pair of numbers measuring the lower and upper bound of the alternative's 'utility interval'. Using these two values, one can then talk about behavior when faced with a particular set of possibilities.[8] Inspired by Fishburn, we use the term vagueness to describe the difference between upper and lower utility of an alternative.

# 3 Nonparametric Identification

Let $(I, \mathcal{F}, \mathbf{P})$ be a non-atomic probability space. We use $i \in I$ to denote a random individual from the population $I$. For each $i \in I$ the set of feasible choices is $\mathcal{A}$, and for each $a \in \mathcal{A}$ decision maker $i$ has a utility interval $[\underline{u}_i(a), \bar{u}_i(a)]$ as described in Section 2.[9] As usual, utility values are

[7]Quoting from Fishburn: "For example, if you prefer your coffee black it seems fair to assume that your preference will not increase as $x$, the number of grains of sugar in your coffee, increases. You might well be indifferent between $x = 0$ and $x = 1$, between $x = 1$ and $x = 2$, ... , but of course will prefer $x = 0$ to $x = 1000$."

[8]Using Fishburn's example, only cups of coffee which contains low amounts of sugar can be chosen. As soon as sugar content is high enough to make a cup of coffee with no sugar at all strictly better, that yields a dominated alternative. That cup, as well as any other that contains more sugar, will not be chosen.

[9]We assume the set of alternatives is the same for all decision makers for simplicity. We explore the role of attention sets in Section 4.4

known to the individual but not to the analyst.

One can treat $\underline{u} : \mathcal{A} \to \mathbb{R}$ and $\bar{u} : \mathcal{A} \to \mathbb{R}$ as two random functions such that for every $a \in \mathcal{A}$, $\mathbf{P}(\underline{u}(a) \leq \bar{u}(a)) = 1$. As in standard discrete choice models, we rule out indifference by assuming that $\underline{u}(a)$ and $\bar{u}(a)$ are continuous random variables for every $a \in \mathcal{A}$. Our objective is to learn about the joint distribution of utilities $\mathcal{U} = \{(\underline{u}(a), \bar{u}(a))\}_{a \in \mathcal{A}}$, or about features of this distribution, using data on choices.

Only the relative position of the lower and upper utilities is choice relevant. Therefore, we are interested in learning about properties of the joint distribution of $\mathcal{U}$ that affect choices. For example, $\mathbf{P}\left(\underline{u}(a) \geq \bar{u}(b)\, \forall b \neq a\right)$ which is the probability that a random individual prefers alternative $a$ to all other alternatives.

Rationality implies that the choice made by a decision maker cannot be dominated by any other alternative. For $i \in I$ let $M_i = M_i(\mathcal{A})$ be the set of alternatives that are not dominated,

$$\begin{aligned} M_i &= \{a \in \mathcal{A} : \nexists b \in \mathcal{A} \text{ such that } \underline{u}_i(b) > \bar{u}_i(a)\} \\ &= \{a \in \mathcal{A} : \max_{b \in \mathcal{A}} \underline{u}_i(b) < \bar{u}_i(a)\}. \end{aligned} \tag{1}$$

We can describe $M$ as a mapping $M : I \to \mathcal{K}(\mathcal{A})$, where $\mathcal{K}(\mathcal{A})$ is the set of all non-empty subsets of $\mathcal{A}$. Since $\mathcal{A}$ is finite, $M_i$ is non empty (a maximum exists) and $\mathcal{K}(\mathcal{A})$ contains compact sets. Since $\underline{u}$ and $\bar{u}$ are random variables, for all $A \in \mathcal{K}(\mathcal{A})$ compact, $\{i : M_i \cap A \neq \emptyset\} \in \mathcal{F}$. Therefore, $M$ is a *random set*. Appendix A provides a summary of the definitions and tools of random set theory that are used in the body of the paper.

For every (non-empty) $A \in \mathcal{K}(\mathcal{A})$ define

$$\theta_A = \mathbf{P}(M = A)$$

to be the proportion of decision makers whose set of non-dominated alternatives is $A$. This definition extends the standard concept of choice probabilities introduced in the model with complete preferences. When preferences are not complete, there is a set $A$ of cardinality bigger than 1 such that $\theta_A > 0$. Complete preferences mean that $\theta_A > 0$ if and only if $|A| = 1$ and $\theta_A = 0$ otherwise (where $|A|$ denotes the cardinality of the set $A$). The collection $\theta = \{\theta_A\}_{A \in \mathcal{K}(\mathcal{A})}$ describes all choice relevant parameters of the joint distribution of $\mathcal{U}$.

The vector of preference parameters $\theta$ satisfies the following properties:

1. $\sum_{A \in \mathcal{K}(\mathcal{A})} \theta_A = 1$. Therefore, after ordering these parameters in some way, the vector of choice parameters is an element of the simplex $\Theta = \Delta(\mathcal{K}(\mathcal{A}))$.[10]

[10] In section 4.3 we discuss abstaining which means that decision makers do not have to choose any alternative in $\mathcal{A}$. Here we assume that the choice probabilities sum to 1 and therefore the choice parameters sum to 1 as well.

2. For any strict subset of parameters $\{\theta_A\}_{\mathcal{K}}$ where $\mathcal{K} \subsetneq \mathcal{K}(\mathcal{A})$ we have $0 \leq \sum_{A \in \mathcal{K}} \theta_A \leq 1$.

Before observing any data, we can only say that the vector of preference parameters lies in the simplex and the sum of any subset of choice parameters lies between 0 and 1 (See Figure 2a in Section 4.1 for an example).

Assumption 1, rationality, implies that individual $i$'s choice, denoted $y_i$, is an element of the random set $M_i$. In the context of random set theory, this behavioral assumption translates to the following measurability assumption.

**Assumption 2 (measurability)** *The random variable $y$ is a selection of the random set $M$, $y \in Sel(M)$.*[11]

A well known result from random set theory, called Artstein's Inequalities, connects the containment functional of the random set $M$ to the distribution function of a selection from that set (see discussion and definitions in Appendix A). Theorem 1 shows there are restrictions that these inequalities impose on features of the joint distribution of preferences and the preference parameters even if preferences are not complete.

**Theorem 1** *Under Assumptions 1 and 2, the identification region associated with the random set $M$ for the choice parameters is given by*

$$\Theta^I = \{\theta \in \Theta : \sum_{A' \subset A} \theta_{A'} \leq \mathbf{P}(y \in A),\ \forall A \subset \mathcal{A}\}. \tag{2}$$

**Proof.** Artstein's Lemma (see Theorem 5 in Appendix A) imply that $y$ is a selection of $M$ if and only if for all $A \subset \mathcal{A}$,

$$C_M(A) \leq \mathbf{P}(y \in A). \tag{3}$$

The choice probabilities on the right-hand side of inequality (3) are identified from the data for any $A \subset \mathcal{A}$. The containment functional (see Definition 6 in Appendix A) on the left-hand side of (3) depends on the unknown joint distribution of $\mathcal{U}$. Therefore, these inequalities impose restrictions on the choice parameters that depend on this unknown distribution. Computing the

[11] The set of selections $Sel(M)$ (Definition 5 in Appendix A) is non-empty, see Theorem 2.13 in Molchanov (2005).

containment functional gives,

$$
\begin{aligned}
C_M(A) &= \mathbf{P}(M \subset A) \\
&= \sum_{A' \subset A} \mathbf{P}(M = A') \\
&= \sum_{A' \subset A} \theta_{A'}.
\end{aligned}
$$

Given our knowledge of $\{\mathbf{P}(y \in K)\}_{K \subset \mathcal{A}}$, the identification set is as defined in equation (2). ■

Artstein's inequalities imply that the set $\Theta^I$ is the sharp identification region for the parameter $\theta$. These inequalities are both sufficient and necessary for a parameter to be included in the identification set and hence $\Theta^I$ is the sharp identification set. This point was made in Beresteanu et al. (2011) and Beresteanu et al. (2012). Therefore, even without completeness, the data imposes restriction on possible parameters of the joint distribution of preferences. The connection between the choice parameters and the order of the upper and lower utilities can be understood from the following relationship.

$$
\begin{aligned}
\mathbf{P}(M \subset A) &= \mathbf{P}(M \cap A^c = \varnothing) \\
&= \mathbf{P}(\max_{k \in A^c} \bar{u}(k) < \max_{a \in \mathcal{A}} \underline{u}(a)) \\
&= \mathbf{P}(\max_{k \in A^c} \bar{u}(k) < \max_{a \in A} \underline{u}(a))
\end{aligned}
$$

Therefore, restrictions imposed on the choice parameters can be translated into restrictions on the order of the upper and lower utilities.

$\Theta^I$ is a strict subset of $\Theta$, if there exists $a \in A$ such that $0 < \mathbf{P}(y = a) < 1$, which is trivially true when there are at least two alternatives. Theorem 1 shows that $\theta_a \leq \mathbf{P}(y = a)$ which is strictly less than 1.[12] The set $\Theta$ does not include this restriction on $\theta_a$ and therefore $\Theta^I \subsetneq \Theta$.

Next, we show that all the inequalities in the definition of the identification region are potentially binding. Let $\mathcal{K}_1$ be a subset of $\mathcal{K}(\mathcal{A})$ that includes all subsets of $\mathcal{A}$ of cardinality 1 and let

$$\Theta^1 = \{\theta \in \Theta : \theta_A = C_M(A) \leq \mathbf{P}(y = a) \, \forall A \in \mathcal{K}_1\}$$

be the set of choice parameters that satisfy Artstein's inequalities for subsets $A \in \mathcal{K}_1$. The following result gives conditions for this set to be larger than the identification set defined in Theorem 1.

[12] To simplify notation, for $A = a$, a singleton, we let $\theta_a = \theta_{\{a\}}$.

**Theorem 2** *If* $\exists A \subset \mathcal{A}$ *such that* $|A| \geq 2$ *and* $\theta_A > 0$*, then* $\Theta^I \subsetneq \Theta^1$*.*

**Proof.** $C_M(\{a\}) = \theta_{\{a\}} \leq \mathbf{P}(y = a) \forall a \in A$, by Theorem 1. Summing over $a \in A$, $\sum_{a \in a} \theta_{\{a\}} \leq \mathbf{P}(y \in A)$. This inequality is part of the set $\Theta^1$. Also, $C_M(A) = \sum_{A' \subset A} \theta_A = \sum_{a \in A} \theta_{\{a\}} + \sum_{A' \subset A : |A'| \geq 2} \theta_A \leq \mathbf{P}(y \in A)$, by Theorem 1. The last inequality implies that $\sum_{a \in A} \theta_{\{a\}} \leq \mathbf{P}(y \in A) - \sum_{A' \subset A : |A'| \geq 2} \theta_A$. This inequality is not part of the definition of $\Theta^1$. By our assumption, $\sum_{A' \subset A : |A'| \geq 2} \theta_A > 0$. Therefore, $\Theta^I$ includes at least one additional binding inequality which is not included in $\Theta^1$. ■

Theorem 2 illustrates that inequalities involving sets with cardinality greater than 1 are potentially binding (in addition to inequalities related to subsets with cardinality 1). In terms of the capacity functional, the condition in Theorem 2 can be replaced with $\exists A \subset \mathcal{A}$ such that $|A| > 2$ and $C_M(A) > \sum_{a \in A} \theta_a$.

**Example:** Suppose the alternatives set is $\mathcal{A} = \{a_0, a_1, a_2\}$. To simplify notation, let the choice parameters be $\theta = (\theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{12}, \theta_{012})$ where $\theta_0 = \theta_{\{a_0\}}$ and similarly for the other subsets. Let $p_j = \mathbf{P}(y = j)$ for $j = 0, 1, 2$ be the choice probabilities. The following set of inequalities have to be satisfied by Theorem 1:

$$
\begin{aligned}
\theta_0 &\leq p_0 \\
\theta_1 &\leq p_1 \\
\theta_2 &\leq p_2 \\
\theta_0 + \theta_1 + \theta_{01} &\leq p_0 + p_1 \\
\theta_0 + \theta_2 + \theta_{02} &\leq p_0 + p_2 \\
\theta_1 + \theta_2 + \theta_{12} &\leq p_1 + p_2 \\
\theta_0 + \theta_1 + \theta_2 + \theta_{01} + \theta_{02} + \theta_{12} + \theta_{012} &\leq p_0 + p_1 + p_2 = 1 \\
\theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{12}, \theta_{012} &\geq 0.
\end{aligned}
$$

The last two inequalities define $\Theta$ (the simplex). The first three inequalities with the last two constitute $\Theta^1$, which uses only inequalities related to sets of cardinality 1. Combining all the inequalities above yields the (sharp) identification region $\Theta^I$ defined in Theorem 1. If either $\theta_{01}$, $\theta_{02}$, $\theta_{12}$ or $\theta_{012}$ are strictly positive, then as Theorem 2 shows, the identification set, $\Theta^I$, is a strict subset of $\Theta^1$. Moreover, it is clear from the above inequalities that the identification set is a convex subset of the simplex and therefore can be easily computed.

So far, we have obtained a general characterization of the identification region, and explored some of its properties. Further results require more restrictions on the environment. For example, in Section 5, we assume the joint distribution of $\mathcal{U}$ belongs to a finite-dimensional parameter family, and focus of partial identification of this parameter. In the next section, however, we avoid parametric assumptions and limit attention to a binary choice; this enables us to illustrate what we have learned so far, as well as present our next main results, using simple pictures and focusing on intuition.

# 4 Binary Choice

In this section we focus on binary choice models. That is, we assume that the set of alternatives is $\mathcal{A} = \{a_0, a_1\}$ for all $i$. As an illustration of Theorem 1, we first derive the identification region for parameters of interest similar to those identified in models with complete preferences, and show that these parameters are only partially identified. We then present a result related to instrumental variables that is the second main result of the paper. One can reduce the identification region and potentially rule out the hypothesis that all individuals have complete preferences. We then study how the identification region is affected by abstention and attention sets.

## 4.1 No-assumptions Bounds

In this Section we derive the identification region implied by Theorem 1, and show how it differs from the point identified case of complete preferences. Since there are only two alternatives, we can illustrate our results with diagrams. We start with the case in which data identifies only the fraction of individuals who chose each alternative. We want to find out what these fractions can tell us about the distribution of utility values in the population. Formally, let $u_i(a_j) = [\underline{\mathrm{u}}_i(a_j), \bar{u}_i(a_j)]$ for $j = 0, 1$; this is the utility interval agent $i$ assigns to alternative $a_j$. Let $\mathcal{U} = (\underline{\mathrm{u}}(a_0), \bar{u}(a_0), \underline{\mathrm{u}}(a_1), \bar{u}(a_1))$ be the corresponding random vector of utilities. We seek to identify, or partially identify, features (parameters) of the joint distribution of $\mathcal{U}$. For binary choice in particular, one is interested in the relative position of the utility intervals $[\underline{\mathrm{u}}(a_0), \bar{u}(a_0)]$ and $[\underline{\mathrm{u}}(a_1), \bar{u}(a_1)]$.

When preferences are complete, each interval is a singleton, and what matters for choice is the order of the utilities resulting from alternatives $a_0$ and $a_1$. Without completeness, an individual chooses from the random set of non-dominated alternatives, $M$, that is defined as

$$M = \begin{cases} \{a_0\} & \text{if } \underline{\mathrm{u}}(a_0) > \bar{u}(a_1) \\ \{a_1\} & \text{if } \underline{\mathrm{u}}(a_1) > \bar{u}(a_0) \\ \{a_0, a_1\} & \text{otherwise.} \end{cases}$$

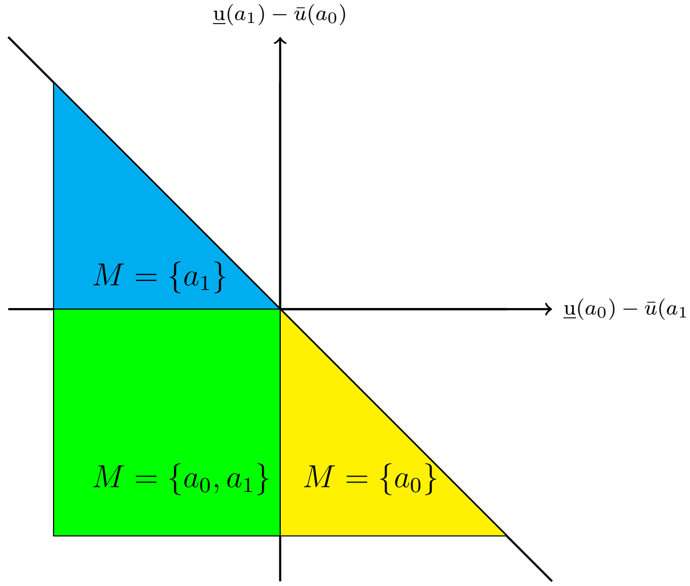



Figure 1: Choice Rule with Incomplete Preferences

Rationality means that $a_1$ is chosen with certainty if $\underline{u}(a_1) > \bar{u}(a_0)$, while $a_0$ is chosen with certainty if $\underline{u}(a_0) > \bar{u}(a_1)$.[13] We display the decision maker's behavior in Figure 1 which illustrates the decision rule in terms of the random set $M$. The axes are given by the difference between the lower utility of one alternative and the upper utility of the other, and the regions illustrating $M$ lie below the $-45°$ line.[14]

We define,

$$\theta_0 = Pr(\underline{u}(a_0) > \bar{u}(a_1))$$
$$\theta_1 = Pr(\underline{u}(a_1) > \bar{u}(a_0))$$

to be the probabilities that a random decision maker prefers alternative $a_0$ over alternative $a_1$ and the probability that a random decision maker prefers alternative $a_1$ over alternative $a_0$, respectively.

If preferences are complete and $\underline{u}(a_j) = \bar{u}(a_j)$ for $j = 0, 1$, then $(\theta_0 + \theta_1) = 1$. Due to incompleteness we can only say that $0 \leq \theta_0 + \theta_1 \leq 1$. Without any data, there are no further restrictions. The identification region of $(\theta_0, \theta_1)$ before observing data is depicted in Figure 2a as the triangle between the axes and the $-45°$ line.

[13] Since we assume $\underline{u}(a_i)$ and $\bar{u}(a_i)$ are continuous random variables equality can be ignored.

[14] By definition, $\bar{u}(a_j) - \underline{u}(a_j) \geq 0$ for $j = 0, 1$; adding over $j$ and rearranging one gets $[\underline{u}(a_1) - \bar{u}(a_0)] \leq -[\underline{u}(a_0) - \bar{u}(a_1)]$. When upper and lower utilities coincide because preferences are complete, then $M$ coincides with the $-45°$ line

Data can help the analyst narrow down possible values of $(\theta_0, \theta_1)$. As usual, aggregate choices are observed by the analyst (while individual utility intervals are not). Let $y_i$ denote the choice made by individual $i$, and denote the choice probability of alternative $a_1$ by $p_1 = \mathbf{P}(y_i = a_1)$ and the choice probability of $a_0$ by $p_0 = \mathbf{P}(y_i = a_0)$. These are the fractions of individuals that choose $a_1$ and $a_0$ respectively, and are identified from the data generating process.

If preferences are complete, one has a point identified model where $(\theta_0, \theta_1) = (p_0, p_1)$. Without completeness, Theorem 1 implies that even if point identification is not possible partial identification is. An agent's choice, $y$, is a selection of the random set $M$ because of Assumption 2. Artstein inequalities imply that $y \in Sel(M)$ if and only if $\mathbf{P}(y \in K) \geq C_M(K)$ for every closed set $K$ where $C_M(K)$ is the containment functional. Substituting $K = \{a_0\}$ and $K = \{a_1\}$, this implies

$$p_0 = Pr(y \in \{a_0\}) \geq C_M(\{a_0\}) = Pr(M \subset \{a_0\}) = Pr(\underline{\mathrm{u}}(a_0) > \bar{u}(a_1)) = \theta_0$$
$$p_1 = Pr(y \in \{a_1\}) \geq C_M(\{a_1\}) = Pr(M \subset \{a_1\}) = Pr(\underline{\mathrm{u}}(a_1) > \bar{u}(a_0)) = \theta_1,$$

and thus the identification region is

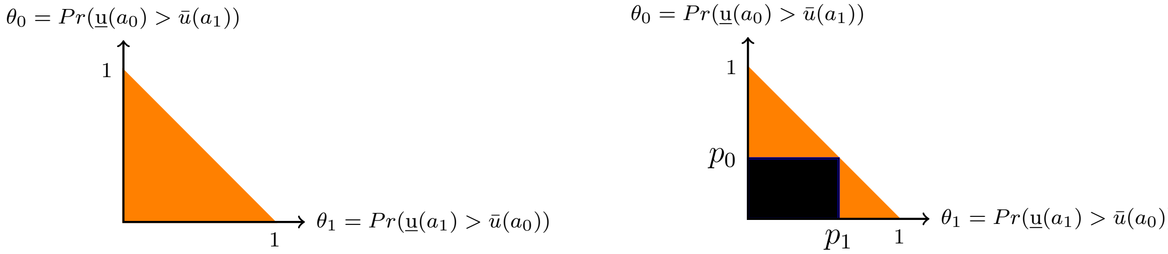
$$\Theta^I = \{(\theta_0, \theta_1) |\ \theta_0 \leq p_0,\ \theta_1 \leq p_1\}. \tag{4}$$

Since all individuals who prefer $a_0$ choose it, and some individuals might choose it even though they cannot compare it with $a_1$, the probability that an individual prefers $a_0$ cannot exceed the fraction $p_0$ of individuals who chooses it. If, for example, $p_0 = 0.37$ then the identification region cannot admit the case where $\theta_0 = 0.4$ and $\theta_1 = 0.6$.

The size of the group of people who choose an outcome even though they cannot compare it to the other could go from zero up to including all those who made that choice. For this reason, the discrete choice model that allows for incomplete preferences is partially identified. The dark area in Figure 2b represents the identification region for $(\theta_0, \theta_1)$ given a pair of $p_0$ and $p_1$ values. The combinations of $(\theta_0, \theta_1)$ which are included in the two light colored triangles in Figure 2b are eliminated from the identification region after observing data. Observing choices informs the researchers about the possible values of the probability that $a_0$ is strictly preferred to $a_1$ and the probability that $a_1$ is strictly preferred to $a_0$.

The identification region in equation (4) can be useful in understanding which features of the theoretical model can be deduced from data. For example, denote with $\nu$ the fraction of decision makers whose preferences are incomplete. Clearly,

$$\nu = 1 - (\theta_0 + \theta_1) \tag{5}$$




(a) Without Data

(b) With Data

Figure 2: Binary Choice

and $\theta_0 + \theta_1$ is the proportion of decision makers who can compare the two alternatives. From Figure 2b one notices that the possibility that $\theta_0 + \theta_1 = 1$ (and thus $\nu = 0$) is included in the identification region: this is the point where the identification region touches the line connecting $\theta_0 = 1$ to $\theta_1 = 1$. Thus, one cannot rule out the possibility that all decision makers have complete preferences. Similarly, one cannot rule out the possibility that all decision makers cannot compare the two alternatives, so that $\theta_0 = \theta_1 = 0$ and $\nu = 1$ (this is the origin). One cannot state anything sharper than $0 \leq \nu \leq 1$ without additional information about preferences.

Suppose one knows that at least a proportion $\underline{\nu} > 0$ of decision makers cannot rank the two alternatives ($\nu \geq \underline{\nu}$).[15] Then $\theta_0 + \theta_1 \leq 1 - \nu \leq 1 - \underline{\nu}$. The corresponding identification region is shown in Figure 3. The dark region includes all points consistent with two statements: at least $\underline{\nu}$ proportion of decision makers cannot rank the two alternatives, and proportion $p_0$ of them chose alternative $a_0$.

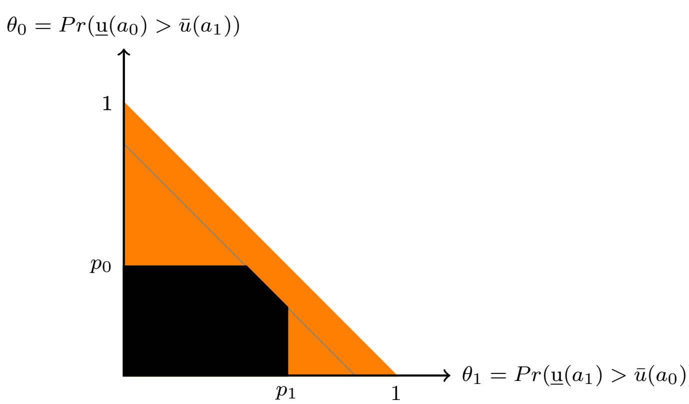



Figure 3: Partial identification with minimal amount of vagueness

Our results so far show that under reasonably few standard assumptions identification, albeit

[15] In the voting data in Section 6 there is information suggesting that at least a certain proportion of voters may not have been able to rank the candidates.

partial, is possible when preferences are not complete. Lack of completeness does not imply that "anything goes". Next, we introduce the concept of an instrumental variable, and show how such a variable could shrink the identification region in an interesting way.

## 4.2 Instrumental variables

In this section we show how instrumental variables can refine the identification region established in Section 4.1. In particular, we define an instrument as a random variable that influences choices while having (almost) no effect on preferences. When these instruments exist, they can be used to rule out the possibility that all decision makers have complete preferences. Intuitively, if observed behavior changes with the realization of this random variable, then it must be the case that some individuals' behavior was not dictated by an actual ranking between the alternatives. In our voting application, a possible instrumental variable is represented by the order in which two candidates are presented on the ballot.

**Perfect Instruments**

Think of a random variable that is independent of utilities but is correlated with choice. Intuitively, whenever a decision maker can rank the alternatives her choice cannot depend of the realized values of this random variable. However, when a decision maker cannot compare the alternatives her choice could depend on the random variable realizations. So, if observed choices change with the realized value of the instrument it must be that some decision makers were not able to rank alternatives. We formalize the idea of an instrument as follows.

**Definition 2** *Let $Z$ be a random variable with a non-empty support $\mathcal{Z}$. For every $z \in \mathcal{Z}$ and for $j = 0, 1$ let $p_{j|z} = \mathbf{P}(y = a_j | Z = z)$. Let,*

$$\Delta_0 = \sup_{z \in \mathcal{Z}} p_{0|z} - \inf_{z \in \mathcal{Z}} p_{0|z}.$$

*We say that $Z$ is an instrumental variable if (1) $(\underline{u}(a_j), \bar{u}(a_j)_{j=0,1})$ are independent of $Z$, and (2) $\Delta_0 > 0$.*

An instrumental variable can then be used to establish the result that some individuals' preferences must not be complete.

**Theorem 3** *Let $Z$ be an instrumental variable. Then,*

$$\theta_0 \leq \inf_{z \in \mathcal{Z}} p_{0|z} \qquad \text{and} \qquad \theta_1 \leq \inf_{z \in \mathcal{Z}} p_{1|z}, \tag{6}$$

*and $\nu > 0$.*

**Proof.** By the independence assumption, $\theta_{0|z} = \mathbf{P}(M = a_0|Z = z) = \mathbf{P}(\underline{u}(a_0) > \bar{u}(a_1)|Z = z) = \mathbf{P}(\underline{u}(a_0) > \bar{u}(a_1)) = \theta_0$ for all $z \in \mathcal{Z}$. Artstein's inequalities applied for each $z \in Z$ as in the proof of Theorem 1 imply that $\theta_0 \leq p_{0|z}$ for every $z \in \mathcal{Z}$. Therefore, $\theta_0 \leq \inf_{z\in\mathcal{Z}} p_{0|z}$ and similarly $\theta_1 \leq \inf_{z\in\mathcal{Z}} p_{1|z}$, establishing equation (6). Using these results and the definition of $\nu$:

$$\nu = 1 - (\theta_0 + \theta_1) \geq 1 - \left(\inf_{z\in\mathcal{Z}} p_{0|z} + \inf_{z\in\mathcal{Z}} p_{1|z}\right) = \sup_{z\in\mathcal{Z}} p_{0|z} - \inf_{z\in\mathcal{Z}} p_{0|z} = \Delta_0 > 0$$

as desired. ■

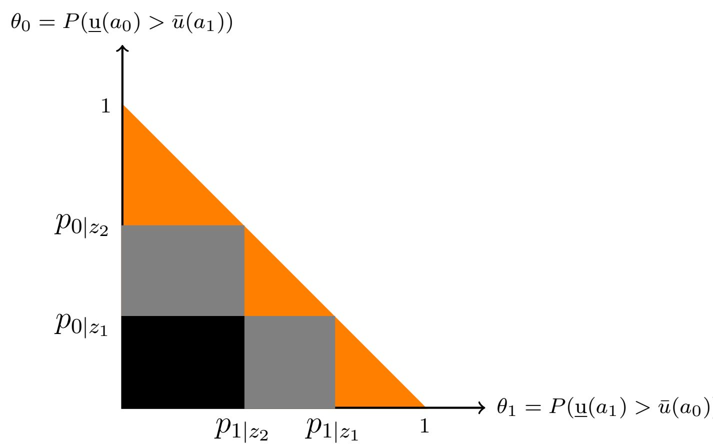



Figure 4: Identification Region with an Instrumental Variable

Theorem 3 says that the binding constraint on the probability that one outcome is preferred to the other is imposed by the lowest conditional probability that outcome is chosen, where conditioning is upon the values of the instrument. Figure 4 illustrates the identification power of having an instrumental variable. The identification region does not contain any point on the $-45°$ line. Therefore, we can rule out the possibility that all individuals had complete preferences. The distance of the identification region from the $-45°$ line depends on $\Delta_0$ which measures the extent to which choices are influenced by $Z$.

**Imperfect Instruments**

The assumption that $Z$ is independent of the distribution of the utilities in $\mathcal{U}$ is a strong assumption that can be relaxed to certain extent explained here. Nevo and Rosen (2012) introduced the notion of imperfect instrumental variables in a linear regression model with endogenous regressors. In their context, an imperfect instrumental variable is a variable $Z$ correlated with the error term of the regression but to a much lesser degree than it is correlated with the endogenous regressor. Nevo and Rosen (2012) show that, under some conditions, imperfect instruments can lead to partial identification of the regression parameters. We adapt the idea of imperfect instrumental variables to our model. Define the following quantities,

$$\theta_{0|z} = \mathbf{P}(\underline{\mathrm{u}}(a_0) > \bar{u}(a_1)|Z = z)$$
$$\theta_{1|z} = \mathbf{P}(\underline{\mathrm{u}}(a_1) > \bar{u}(a_0)|Z = z),$$

and

$$\delta_0 = \theta_0 - \inf_{z\in\mathcal{Z}} \theta_{0|z}$$
$$\delta_1 = \theta_1 - \inf_{z\in\mathcal{Z}} \theta_{1|z}.$$

When Z is independent of preference parameters, as in Definition 2, $\delta_0 = \delta_1 = 0$ and one has a perfect instrument. If $Z$ is not a perfect instrument, $\delta_0$ and $\delta_1$ measure the sensitivity of the utilities to changes in the value of the instrument $Z$.

**Definition 3** *Let $Z$ be a random variable with a non-empty support $\mathcal{Z}$. For every $z \in \mathcal{Z}$ and for $j = 0, 1$ let $p_{j|z} = \mathbf{P}(y = a_j|Z = z)$. We say $Z$ is an imperfect instrumental variable if $\Delta_0 > \delta_0 + \delta_1$.*

The next result shows that if utilities depend on $Z$ to a lesser extent than choices depend on $Z$, the fraction of decision makers whose preferences are not complete is bounded away from 0 by a strictly positive quantity.

**Theorem 4** *If $Z$ is an imperfect instrumental variable, then $\nu > 0$.*

**Proof.** As before, Artestein's inequalities imply that $\theta_{0|z} \leq p_{0|z}$ for all $z \in Z$; and therefore

$$\inf_{z\in\mathcal{Z}} \theta_{0|z} \leq \inf_{z\in\mathcal{Z}} p_{0|z}$$

Using the definition of $\delta_0$

$$\theta_0 - \delta_0 \leq \inf_{z\in\mathcal{Z}} p_{0|z}.$$

and similarly $\theta_1 - \delta_1 \leq \inf_{z\in\mathcal{Z}} p_{1|z}$. Summing these two inequalities gives

$$(\theta_0 + \theta_1) - (\delta_0 + \delta_1) \leq \inf_{z\in\mathcal{Z}} p_{0|z} + \inf_{z\in\mathcal{Z}} p_{1|z}.$$

Since $\inf_{z\in\mathcal{Z}} p_{0|z} = 1 - \sup_{z\in\mathcal{Z}} p_{0|z}$, we can write,

$$(\theta_0 + \theta_1) - (\delta_0 + \delta_1) \leq 1 - \Delta_0.$$

Finally,

$$\nu = 1 - (\theta_0 + \theta_1) \geq \Delta_0 - (\delta_0 + \delta_1) > 0.$$

■

### 4.3 Abstention

In this Section, we consider the possibility that choices are unobserved for a subset of decision makers. This situation can occur for several reasons. First, as is the case with voting (see Section 6), individuals can refrain from making a decision (abstention). Second, the data collected is incomplete due to non-response or non-observability. Let $p_0$, $p_1$, and $\gamma$ represent the proportion of decision makers who chose alternative $a_0$, alternative $a_1$ and abstained, respectively, such that $p_0 + p_1 + \gamma = 1$. Let $V \in \{0, 1\}$ be a binary random variable indicating whether an individual's choice is observed, $V = 1$, or unobserved, $V = 0$.

**No assumptions bounds**

To identify the probability that a random individual strictly prefers $a_0$ over $a_1$ given that this decision maker's choice is observed one can use the methods described in section 4.1. For example, one could assume that voters that went to the polls and then abstained must have done so because their preferences were incomplete. However, we are interested in identification when such an assumption is not made. In particular, we want to identify $\theta_0$ and $\theta_1$ in the entire population of decision makers, not only among those who made a choice. Using the total law of probability (see Chapter 1 in Manski (2003) and references therein for first applications of the law of total probability in partial identification), we can write

$$\begin{aligned} \theta_0 &= \mathbf{P}(\underline{\mathrm{u}}_0 > \bar{u}_1) = (1-\gamma)Pr(\underline{\mathrm{u}}_0 > \bar{u}_1 | V = 1) + \gamma Pr(\underline{\mathrm{u}}_0 > \bar{u}_1 | V = 0), \\ \theta_1 &= \mathbf{P}(\underline{\mathrm{u}}_1 > \bar{u}_0) = (1-\gamma)Pr(\underline{\mathrm{u}}_1 > \bar{u}_0 | V = 1) + \gamma Pr(\underline{\mathrm{u}}_1 > \bar{u}_0 | V = 0). \end{aligned} \tag{7}$$

Rationality implies that

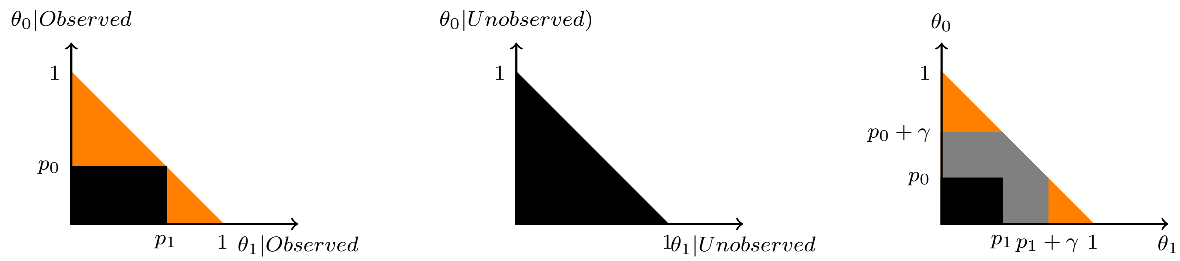
$$\begin{aligned} \mathbf{P}(\underline{\mathrm{u}}_0 > \bar{u}_1 | V = 1) &\leq p_0, \\ \mathbf{P}(\underline{\mathrm{u}}_1 > \bar{u}_0 | V = 1) &\leq p_1. \end{aligned} \tag{8}$$

These inequalities are described by the rectangular identification region in Figure 5a. We denote this set as $\Theta^O = \{(\theta_0, \theta_1) : 0 \leq \theta_0 \leq p_0 \,,\, 0 \leq \theta_1 \leq p_1\}$.

The quantities $\mathbf{P}(\underline{\mathrm{u}}_0 > \bar{u}_1 | V = 0)$ and $\mathbf{P}(\underline{\mathrm{u}}_1 > \bar{u}_0 | V = 0)$ are unidentified and satisfy the following inequality

$$0 \leq Pr(\underline{\mathrm{u}}_0 > \bar{u}_1 | V = 0) + Pr(\underline{\mathrm{u}}_1 > \bar{u}_0 | V = 0) \leq 1.$$

Denote this set as $\Theta^U = \{(\theta_0, \theta_1) : 0 \le \theta_0 + \theta_1 \le 1\}$. These inequalities are represented by the triangular identification region in Figure 5b.




(a) Observed (b) Unobserved (c) Joint

Figure 5: Partial identification with $\gamma$ abstaining

Finally, we can combine both identification regions $\Theta^O$ and $\Theta^U$ using equation (7),

$$\Theta^I = (1-\gamma)\Theta^O \oplus \gamma\Theta^U, \tag{9}$$

where $\oplus$ is the Minkowski sum. $\Theta^I$ is the small rectangular region in Figure 5c.

**Instrumental variables.**

One can combine these ideas with the notion of instrumental variables. Let $Z$ be an instrumental variable with a finite support $\mathcal{Z}$ satisfying Definition 2. Equation (7) holds for all values of the instrument $Z$. Furthermore, we assume that $\gamma$, the probability of abstaining remains constant across different values of the instrument $Z$. Equation (8) holds for all values of the instruments which affect only the right hand-side. Therefore,

$$\begin{aligned} \mathbf{P}(\underline{u}_0 > \bar{u}_1 | V=1) &\le \inf_{z\in\mathcal{Z}} p_{0|Z=z}, \\ \mathbf{P}(\underline{u}_1 > \bar{u}_0 | V=1) &\le \inf_{z\in\mathcal{Z}} p_{1|Z=z}. \end{aligned} \tag{10}$$

Combining equation (7) and equation (10), we have

$$\begin{aligned} \theta_0 &\le (1-\gamma)\inf_{z\in\mathcal{Z}} p_{0|Z=z} + \gamma\theta_{0|A=Unobs}, \\ \theta_1 &\le (1-\gamma)\inf_{z\in\mathcal{Z}} p_{1|Z=z} + \gamma\theta_{1|A=Unobs}. \end{aligned} \tag{11}$$

Combining this information with the fact that $0 \le \theta_{0|V=0} + \theta_{1|V=0} \le 1$, we can write the joint identification region similarly to equation (7),

$$\Theta^{I|Z} = (1-\gamma)\Theta^{O|Z} \oplus \gamma\Theta^U, \tag{12}$$

where $\Theta^{O|Z} = \{(\theta_0, \theta_1) : \theta_0 \le \inf_{z\in\mathcal{Z}} p_{0|Z=z}. \theta_1 \le \inf_{z\in\mathcal{Z}} p_{1|Z=z}\}$. Theorem 3 implies that if $\sup_{z\in\mathcal{Z}} p_{0|z} - \inf_{z\in\mathcal{Z}} p_{0|z} > 0$, then $\Theta^{O|Z} \subsetneq \Theta^O$ and therefore $\Theta^{I|Z} \subsetneq \Theta^I$.

**Discussion**

The literature on partial identification can be divided into two subgroups based on the source of partial identification.

1. **Imperfect data:** The data generating process provides observations which are imperfect in some way. Examples include, missing observations due to non-response or attrition, interval valued observations, unobserved couterfactuals etc. The model employed by the researcher would provide a unique prediction of the joint distribution of the observables should the data generating process provide perfect data.

2. **Incomplete models:** The behavioral and mathematical/statistical assumption are not "tight" enough to lead to a unique prediction of the joint distribution of the observables. A leading example is multiple equilibria in games and other models of strategic interaction as well the model of incomplete preferences in this paper.

In both cases sample size is not the issue and the cause of partial identification will persist even when infinite number of observations are provided. The identification regions reported in equations (9) and (12) are a convex combination of both sources of partial identification. $\Theta^U$ is the result of decisions makers whose choice is unobserved (e.g. voters abstaining). $\Theta^O$ and $\Theta^{O|Z}$ are a result of an incomplete model. The resulting identification regions $\Theta^I$ and $\Theta^{I|Z}$ are a convex combination of both sources of partial identification using the weight $\gamma$.

## 4.4 Attention Sets

In Section 4.2 we show that instrumental variables can allow the analyst to provide evidence that at least some decision makers have preferences that are not complete. The origin $(0,0)$ is included in all identification regions presented so far (see Figure 4). Thus, one cannot rule out the possibility that all decision makers are unable to rank the alternatives. In this section we explore a simple assumption that allows the analyst to exclude this possibility.

Consider the case where a known proportion of the decision makers considers, or pays attention to, only a subset of alternatives $\mathcal{A}' \subsetneq \mathcal{A}$. This could happen because some agents are unaware an alternative exist, or because some agents would never consider an alternative even when they are aware of it (for example, members of a political party would never vote for candidates who belong to a different party). When this happens, the set of non-dominated alternatives in Definition (1) is applied to the decision maker with attention set $\mathcal{A}'$ and gives the random set of non-dominated alternatives $M(\mathcal{A}')$.

In a binary choice context $\mathcal{A} = \{a_0, a_1\}$, and there are only three potential attention sets: $\mathcal{A} = \{a_0, a_1\}$, $\mathcal{A}_0 = \{a_0\}$, and $\mathcal{A}_1 = \{a_1\}$. Let $\pi_0$ be the fraction of decision makers who have attention set $\mathcal{A}_0$ and $\pi_1$ the fraction that have attention set $\mathcal{A}_1$, and assume that $\pi_0$ and $\pi_1$ are known to the analyst. Since decision makers with attention set $\mathcal{A}_0$ do not consider, or are unaware of, alternative $a_1$ their choice of $a_0$ is deterministic from the analyst's standpoint. Similar reasoning applies to decision makers with attention set $\mathcal{A}_1$. This model is coherent if $\pi_0 \leq p_0$ and $\pi_1 \leq p_1$. In this case

$$\pi_0 \leq \theta_0 \leq p_0,$$
$$\pi_1 \leq \theta_1 \leq p_1.$$

If either $\pi_0$ or $\pi_1$ are strictly positive the identification set does not contain the origin. The case where $\pi_0, \pi_1 > 0$ is illustrated in figure 6 where the identification set of $(\theta_0, \theta_1)$ is the darker rectangle. The combination of attention sets and instrumental variables assumptions is demonstrated using our empirical application in Section 6.

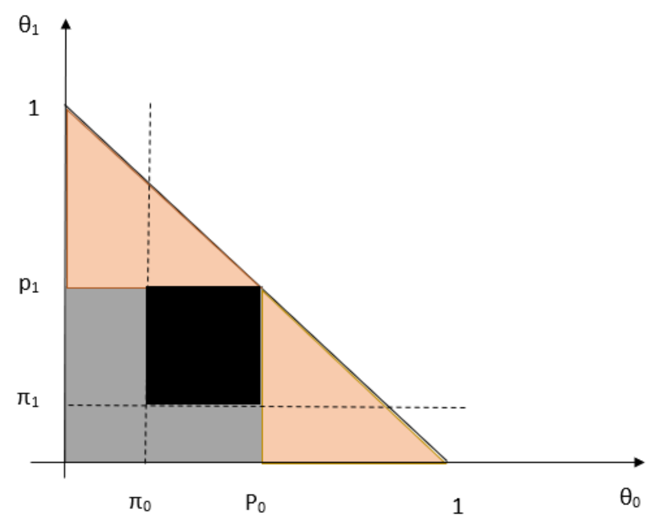



Figure 6: Partial identification with attention sets

# 5 Extensions

In this section we illustrate how some of our identification results can be adapted to additional assumptions about individuals' behavior or the data generating process. In particular, we discuss a version of the model in which upper and lower utility depend linearly on some characteristics (Section 5.1), incompleteness due to ambiguity (Section 5.2) and minmax regret behavior (Section 5.3). For ease of exposition, we focus on binary choice throughout.

## 5.1 Incomplete Discrete Choice Model

We turn our attention to parametric models of discrete choice. Following section 2.1, one can think of the utility of individual $i$ as an interval. In particular, let these utilities be as follows:

$$\begin{aligned}\underline{u}_i(a_0) &= \beta_{0i} - \varepsilon_{0i}\\ \bar{u}_i(a_0) &= \beta_{0i} + \sigma_{0i} - \varepsilon_{0i}\\ \underline{u}_i(a_1) &= \beta_{1i} - \varepsilon_{1i}\\ \bar{u}_i(a_1) &= \beta_{1i} + \sigma_{1i} - \varepsilon_{1i}.\end{aligned}$$

We can subtract $\beta_{0i} - \varepsilon_{i0}$ from all four utilities and define $\beta_i = \beta_{1i} - \beta_{0i}$ and $\varepsilon_i = \varepsilon_{1i} - \varepsilon_{i0}$, so that the normalized upper and lower utilities for decision maker $i$ are defined as

$$\begin{aligned}\underline{u}_{i0} &= 0\\ \bar{u}_{i0} &= \sigma_{0i}\\ \underline{u}_{i1} &= \beta_i - \varepsilon_i\\ \bar{u}_{i1} &= \beta_i + \sigma_{1i} - \varepsilon_i.\end{aligned} \tag{13}$$

We assume that $\varepsilon_i$'s are independent and identically distributed across decision makers and $Var(\varepsilon_i) = 1$. Subtracting $\beta_0 - \varepsilon_{0i}$ from all utilities (location normalization) and assuming unit variance for $\varepsilon_i$ (scale normalization) is necessary to (partially) identify the parameters of the model in equation (13). In addition to location and scale normalization above, one can also assume that $\varepsilon \sim F_\varepsilon$ is continuously distributed. For example, assuming that $\varepsilon_{0i}$ and $\varepsilon_{1i}$ are independently distributed Type 1 Extreme Value implies that $\varepsilon_i$ has a Logistic distribution. This assumption makes the above model an extension of the Logit model.

In most applications of discrete choice one assumes that $\beta_i = X_i\beta$ where $X_i$ is the vector of observed characteristics for individual $i$ or of alternative 1 as perceived by individual $i$. For simplicity, suppose that there are no individual specific characteristics and $\beta_i = \beta$ is constant across all decision makers. Moreover, assume that $\sigma_{i0} = \sigma_{i1} = \sigma \geq 0$ is constant across both choices and all decision makers. In other words, there is only one source of heterogeneity in this model - the unobserved characteristic $\varepsilon_i$.

Following the discussion in section 4.1, the identification region can be written as

$$p_0 \geq \theta_0 = \mathbf{P}(\underline{\mathrm{u}}_0 > \bar{u}_1) = \mathbf{P}(\varepsilon_i > \beta + \sigma)$$
$$p_1 \geq \theta_1 = \mathbf{P}(\underline{\mathrm{u}}_1 > \bar{u}_0) = \mathbf{P}(\varepsilon_i < \beta - \sigma)$$

Combining these inequalities gives,

$$\mathbf{P}(\varepsilon_i < \beta - \sigma) \leq p_1 \leq \mathbf{P}(\varepsilon_i \leq \beta + \sigma).$$

If one assumes that $\varepsilon$ has a continuous everywhere monotone CDF $F_\varepsilon$, the identification set is

$$\Theta^I = \left\{(\beta, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \beta - \sigma \leq F_\varepsilon^{-1}(p_1) \leq \beta + \sigma\right\}.$$

Figure 7 shows the joint identification region for $(\beta, \sigma)$. The point $(0, F_\varepsilon^{-1}(p_1))$ is included in the identification region in Figure 7. This point corresponds to a situation where all decision makers can perfectly rank the two alternatives.

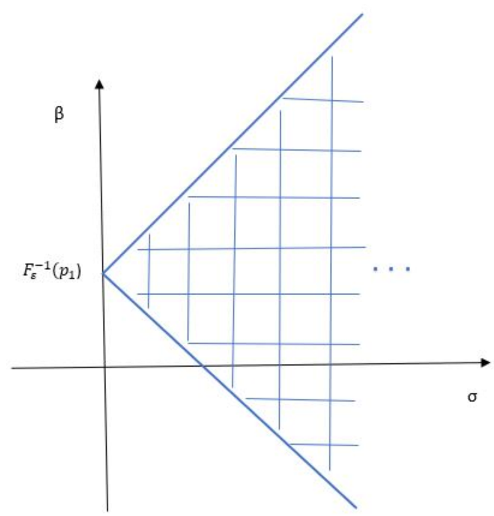



Figure 7: One Dimensional Heterogeneity

If there exist a random variable $Z$ such that the distribution of $\varepsilon|Z$ changes over the support of $Z$ while the determinants of utility, $\beta$ and $\sigma$, remain constant across the values of $Z$, the corresponding identification region is

$$\Theta^I = \left\{(\beta, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \beta - \sigma \leq \inf_{z \in \mathcal{Z}} F_\varepsilon^{-1}(p_0|z) \text{ and } \sup_{z \in \mathcal{Z}} F_\varepsilon^{-1}(p_0|z) \leq \beta + \sigma\right\}.$$

One can use an instrumental variable to reject the hypothesis that upper and lower utilities coincide. When $\varepsilon$ has a Logistic distribution, one gets $F^{-1}(p_1) = \log\left(\frac{p_1}{p_0}\right)$.

## 5.2 Discrete Choice With Knightian Uncertainty

Our identification framework can be modified to allow for different decision theoretic models of behavior when preferences are not complete as long as behavior is described by two numbers. In the following, we illustrate how this could be done for Knightian uncertainty as described in Bewley (1986).[16] Bewely shows that a strict preference relation that is not necessarily complete, but satisfies all other axioms of the standard Anscombe-Aumann framework, can be represented by a family of expected utility functions generated by a unique utility index and a set of probability distributions. Lack of completeness is thus reflected in multiplicity of beliefs: the unique subjective probability distribution of the standard expected utility framework is replaced by a set of probability distributions. When the preference relation is complete this set becomes a singleton.

Let $S$ denote the state space and, with abuse of notation, its cardinality. $\Delta(S)$ is the set of all probability distributions over $S$. Given an alternative $x \in X \subset \boldsymbol{R}^{\boldsymbol{S}}$, $u(x(s))$ denotes the utility that alternative yields in state $s$. If $\pi \in \Delta(S)$, the expected utility of individual $i$ according to $\pi$ is given by

$$E_\pi[u(x)] \equiv \sum_{s \in S} \pi(s) u(x, s).$$

Let $\Pi \subset \Delta(S)$ be a closed and convex set of probability distributions on $S$. According to Bewley's Knightian Decision Theory, decision maker $i$'s preferences $\succ$ are described by the following result:

$$x \succ y \qquad \text{if and only if} \qquad E_\pi[u(x)] > E_\pi[u(y)] \text{ for all } \pi \in \Pi. \tag{14}$$

If the inequality in (14) changes direction for different probability distributions in $\Pi$, the two alternatives are not comparable.

This model is not a special case of the interval order presented previously because comparisons between two alternatives are made one probability distribution at the time. Even though the values of $E_\pi[u(\cdot)]$ form an interval, comparisons are not made looking at the extremes of that interval. Two alternatives could be ranked even if the corresponding utility intervals overlap. Despite this difference, the results of the previous section can be applied here by taking advantage of some simple algebra. Rearranging equation (14) one gets

$$x \succ y \qquad \text{if and only if} \qquad E_\pi[u(x) - u(y)] > 0 \text{ for all } \pi \in \Pi$$

and therefore

$$x \succ y \qquad \text{if and only if} \qquad \min_{\pi \in \Pi} E_\pi[u(x) - u(y)] > 0$$

[16] Bewley's original paper has been published recently as Bewley (2002).

We can then adapt the definition of the set of non-dominated alternatives as follows

$$M = \begin{cases} \{a_0\} & \text{if } \min_{\pi\in\Pi} E_\pi[u(a_0) - u(a_1)] > 0 \\ \{a_1\} & \text{if } \min_{\pi\in\Pi} E_\pi[u(a_1) - u(a_0)] > 0 \\ \{a_0, a_1\} & \text{otherwise.} \end{cases}$$

As before, we let $\theta = (\theta_0, \theta_1)$ be the probabilities that alternative $a_0$ is preferred to $a_1$ and the probability that alternative $a_1$ ir preferred to $a_0$, respectively. By definition,

$$\theta_0 = \mathbf{P}(\min_{\pi\in\Pi} E_\pi[u(a_0) - u(a_1)] > 0)$$
$$\theta_1 = \mathbf{P}(\min_{\pi\in\Pi} E_\pi[u(a_1) - u(a_0)] > 0)$$

From here on, the analysis can proceed along lines similar to the ones provided in Section 4, and we thus leave the details to the reader.

## 5.3 Minmax regret

So far, we have been silent about the way individuals make their choices when two or more alternatives are not comparable. This approach resulted in partial identification of the parameters of interest. Here we focus on the idea of minimizing maximal regret as a way to break the consumers' indecision.

Assume that after a decision is made the individual learns which of the possible utility values represent her 'true' utility for each alternative. Regret occurs if the realized utility of the chosen option turns out to be lower than the realized utility of an alternative that was not chosen.

For a binary choice situation, recall that

$$M = \begin{cases} \{a_0\} & \text{if } \underline{u}(a_0) > \bar{u}(a_1) \\ \{a_1\} & \text{if } \underline{u}(a_1) > \bar{u}(a_0) \\ \{a_0, a_1\} & \text{otherwise.} \end{cases}$$

and thus when the two utility intervals overlap both choices can be rational. If the individual chooses $a_0$ her maximal regret is $\bar{u}(a_1) - \underline{u}(a_0)$, while if she chooses $a_1$ her maximal regret is $\bar{u}(a_0) - \underline{u}(a_1)$. Therefore, if the decision maker minimizes her maximal regret when undecided, the corresponding rational choice region is as follows:

$$M^{minmax} = \begin{cases} \{a_0\} & \text{if} \quad \underline{u}(a_0) - \bar{u}(a_1) > 0 \quad \text{or} \quad \underline{u}(a_0) - \bar{u}(a_1) > \underline{u}(a_1) - \bar{u}(a_0) \\ \{a_1\} & \text{if} \quad \underline{u}(a_1) - \bar{u}(a_0) > 0 \quad \text{or} \quad \underline{u}(a_0) - \bar{u}(a_1) < \underline{u}(a_1) - \bar{u}(a_0) \end{cases} \tag{15}$$

The minmax regret rule selects one alternative from the random set $M$. Figure 8 describes the choice rule in equation (15). The region in Figure 1 where choice was indeterminate is now split between alternatives so that points above the 45 degree line mean alternative $a_1$ is chosen and points below it mean alternative $a_0$ is chosen. Since choice is no longer indeterminate, the corresponding model is point identified.

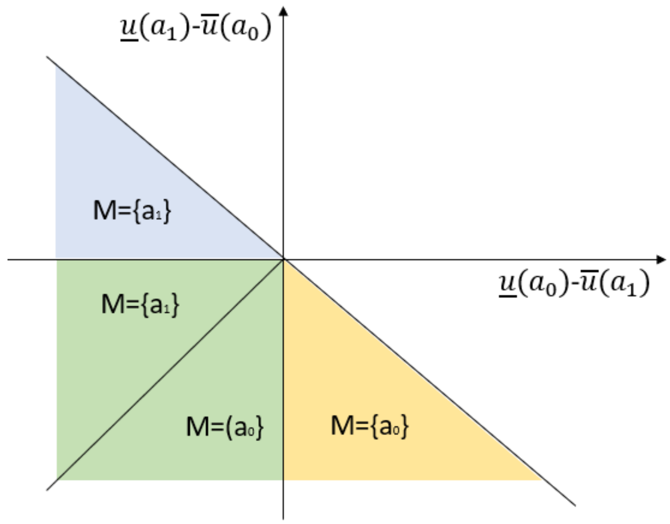



Figure 8: Choice rule with ex-post regret

# 6 Empirical Example - Elections

We implement the methods described in previous sections using data on voting.[17] Specifically, we use precinct-level data obtained from the Ohio Board of Elections for the 2018 midterm elections in Lorain county, and focus on the two races for Justice of the Supreme Court of Ohio. We focus on one county because it consists of a relatively homogeneous population and the covariates used in subsections 6.1 provide decent explanation power. Election data fits our framework in many respects, the main one being that individuals' choices are not observable and one can only observe the share received by each candidate. The data from Ohio is also interesting because candidates' order on the ballot varies from precinct to precinct. This change of order provides a good example of an instrumental variable. A full description of the data on Ohio 2018 midterm elections is in Appendix B including an explanation how candidates' order on the ballot rotates.

In the 2018 midterm elections voters in Ohio participated in eight statewide races. In six of these races candidates' party affiliation was listed on the ballot, while in the remaining two it was not. The races where no party affiliation was noted on the ballot were the two races for the position of Justice of the Supreme Court of Ohio, and therefore we focus on those. Candidates had been selected in party-run primaries, and thus were affiliated with a party, but their affiliation was not actually printed on the ballot itself. As Table 1 shows, in Lorain county the percentage

[17]Since this paper deals with identification, we treat estimators as population lever quantities and leave the statistical issues for future research.

of voters who refrained from expressing a preference is above 20% when party affiliation is not listed, with about 40,000 voters who went to the polls, voted in most state-wide races, but did not vote in the two state-wide races where a candidate's party affiliation was not printed on the ballot.[18] This is an example of what is sometimes called roll-off voting: fewer votes are cast in down-ballot races. It represents a particular form of abstention because the voter has already incurred the costs associated with going to the polls.

| | Total Voters | % of Turnout |
|---|---|---|
| Turnout | 116,231 | 100.00 |
| Governor and Lieutenant Governor | 114,558 | 98.56 |
| Attorney General | 110,236 | 94.84 |
| Auditor of State | 111,021 | 95.52 |
| Secretary of State | 111,994 | 96.35 |
| Treasurer of State | 110,919 | 95.43 |
| U.S. Senator | 113,876 | 97.97 |
| Justice of the Supreme Court 1 | 87,525 | 75.30 |
| Justice of the Supreme Court 2 | 85,472 | 73.54 |

Table 1: Lorain County Statewide Races

The recent literature on ballot roll-off in judicial elections (see Hall and Bonneau (2008) or Marble (2017)), is mostly focused on empirically measuring it and its possible determinants. Similarly to what we find, this literature shows that affiliation on the ballot decreases ballot roll-off. There is also a theoretical literature explaining that abstention could stem from asymmetric information (Feddersen and Pesendorfer (1999)), or context-dependent voting (Callander and Wilson (2006)). In our framework, there could be an alternative reason for roll-off voting: incomplete preferences. When unable to rank alternatives, voters may decide not to make a choice and therefore do not vote in the corresponding race. In the extreme case in which all those who came to the polls and did not vote behaved that way because they could not compare the candidates we would conclude that about 20% of the electorate could not compare the two candidates. In what follows, we will disregard this possibility to illustrate the methods developed in Section 4.

The two races for a seat on the Ohio Supreme Court were Baldwin versus Donnelly and DeGenaro versus Stewart. As mentioned above, for these races the party affiliation of the candidates was not indicated on the ballot. We know, however, that Donnely, in the first race, and Stewart,

[18]Similar numbers hold statewide, with more than 800,000 voters not voting in the races for Justice of the Supreme Court after having gone to the polls.

in the second race, were affiliated with the Democratic Party. Table 2 describes the results for these two races in Lorain county.

| | Baldwin votes | Donnelly votes | Total |
|---|---|---|---|
| Justice of the Supreme Court 1 | 29,564<br>33.8% | 57,961<br>66.2% | 87,525 |
| | DeGenaro votes | Stewart votes | Total |
| Justice of the Supreme Court 2 | 37,282<br>43.6% | 48,190<br>56.4% | 85,472 |

Table 2: Ohio Supreme Court Races - Lorain County Results

## 6.1 Conditional Choice Probabilities

So far, we have focused on identification regions described by inequalities based on unconditional probabilities. Using the results in Table 2, identification regions similar to the one presented in Figure 2b can be constructed for the Supreme Court races. It is clear, however, that the precincts in Lorain county contain a diverse set of voters, and the choice probabilities in these races varied widely from precinct to precinct. For example, Donnelly's (unweighted) average at the precinct level was 60.4% with a standard deviation of 10.5% and Stewart averaged 50.4% at the precinct level with a standard deviation of 12.5%. In this case, one needs to estimate conditional choice probabilities that control for covariates that may impact both choice probabilities and utilities.

Suppose there is a vector of random variables $X$ affecting the utility from each alternative and therefore the choice made by a decision maker.[19] In the context of voting, $X$ can represent the political views of the decision makers in each precinct and how far these are from the views of each candidate. In the case of two alternatives, $\mathcal{A} = \{a_0, a_1\}$, the inequalities in equation (3) hold for each value of the covariate $X$,

$$C_M(K|X = x) = \mathbf{P}(M \subset K|X = x) \leq \mathbf{P}(y \in K|X = x) = p_{y|x}. \tag{16}$$

Conditional choice probabilities, capacity functionals, and the bounds in equation (16) can be estimated separately for each precinct. Doing so effectively conditions on the covariates that characterize the voters in each location in a non-parametric way. Alternatively, we assume a parametric form for the conditional choice probabilities, $Pr(y \in K|X = x)$ and use information from a set of precincts in one county. Specifically, we assume that the conditional choice

[19] See Beresteanu et al. (2012), Section 2 for exact conditions on the probability space and covariates which enable considering conditional probabilities.

probabilities follow a logistic functional form so that for precinct $j$ we have the following:

$$p_{a_1|X=x_j} = \frac{exp(x_j'\beta + \xi_j)}{1 + exp(x_j'\beta + \xi_j)}, \tag{17}$$

where $\xi_j$ is an unobserved characteristic of precinct $j$. Using the inversion suggested by Berry (1994), we can write

$$\log\left(\frac{p_{a_1|X=x_j}}{p_{a_0|X=x_j}}\right) = x_j'\beta + \xi_j. \tag{18}$$

Lorain county includes 191 precincts. For each precinct we observe the percent of registered voters who are registered Democrats, Republicans or Independent, the percent of registered voters who came to vote (turnout) and the average age of registered voters. We use these variables as covariates $x_j$ in Equation (18). The estimators of the coefficients $\beta$ and of the marginal effects of the covariates are reported in Table 3. The predicted choice probabilities calculated at the mean values of the covariates are

$$P(y = Donnelly|\overline{Dem}, \overline{Age}) = 0.668$$
$$P(y = Stewart|\overline{Dem}, \overline{Age}) = 0.576,$$

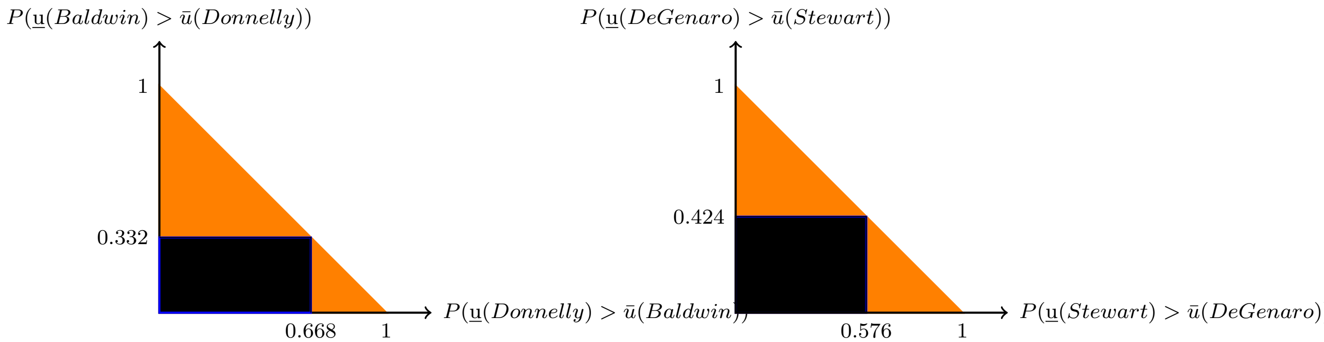
where $\overline{Dem} = 0.207$ and $\overline{Age} = 51.56$.

| Baldwin vs. Donnelly: | | |
|---|---|---|
| variable | coefficient | Marginal effect |
| Intercept | 2.08 (0.363) | - |
| Registered Democrat | 2.66 (0.186) | 0.59% |
| Age | $-0.036$ (0.006) | -0.81% |
| $N = 191$ $R^2 = 0.598$ | | |

| DeGenaro vs. Stewart: | | |
|---|---|---|
| variable | coefficient | Marginal effect |
| Intercept | 2.51 (0.345) | - |
| Registered Democrat | 3.39 (0.279) | 0.84% |
| Age | $-0.056$ (0.007) | -1.37% |
| $N = 191$ $R^2 = 0.665$ | | |

Table 3: Logistic Function Estimation for CCPs
Two Ohio Supreme Court Races

Using these values, one obtains the identification regions in Figures 9a and 9b.




(a) Baldwin vs. Donnelly

(b) DeGenaro vs. Stewart

Figure 9: Ohio's Supreme Court Races - 2018 Midterm Elections
No Assumptions Bounds

## 6.2 Candidates Order

Political science literature suggests that the order in which candidates are presented on the ballot may affect the chances of these candidates to be elected (see, for example, Krosnick et al. (2004) and Meredith and Salant (2013)). As a result, ordering the candidates in alphabetical order may favor candidates with certain family names. The Ohio Board of Elections rotates the order in which candidates appear on the ballot among the precincts (see appendix B). As a result, a candidate may appear first on the ballot in one precinct and last in another nearby precinct. Therefore, by construction, the order of a candidate on the ballot presents an example of an instrumental variable. We focus again on Lorain county to achieve a relatively homogeneous population given the covariates we use. We create a dummy variable indicating whether a candidate appeared first on the ballot at a certain precinct.

$$
\begin{aligned}
P(y = Donnelly|\overline{Dem}, \overline{Age}, first = 0) = 0.652 \\
P(y = Donnelly|\overline{Dem}, \overline{Age}, first = 1) = 0.684 \\
P(y = Stewart|\overline{Dem}, \overline{Age}, first = 0) = 0.556 \\
P(y = Stewart|\overline{Dem}, \overline{Age}, first = 1) = 0.597
\end{aligned}
$$

As one can see from Table 4, in both races for Ohio's supreme court judgeship, being first on the ballot gives a certain advantage over being second.[20] In the Baldwin versus Donnelly

[20] The advantage of appearing first on the ballot, when averaged over the whole state, is rather small. In some counties the effect of the order is rather small, statistically insignificant or even negative. This uncounted difference between counties may be due to omitted covariates. We solve this issue by focusing on one county. Further empirical investigation is left for future work.

| Baldwin vs. Donnelly: | | |
|---|---|---|
| variable | coefficient | Marginal effect |
| Intercept | 2.01 (0.354) | - |
| Registered Democrat | 2.66 (0.161) | 0.59% |
| Turnout Percent | 0.102 (0.170) | -0.022% |
| Age | $-0.039$ (0.0084) | -0.86% |
| first | 0.141 (0.027) | 3.11% |
| $N = 191$ $R^2 = 0.651$ | | |

| DeGenaro vs. Stewart: | | |
|---|---|---|
| variable | coefficient | Marginal effect |
| Intercept | 2.15 (0.467) | - |
| Registered Democrat | 3.43 (0.291) | 0.84% |
| Turnout Percent | $-0.668$ (0.223) | -0.16% |
| Age | $-0.044$ (0.011) | -1.08 % |
| first | 0.176 (0.031) | 4.28% |
| $N = 191$ $R^2 = 0.732$ | | |

Table 4: Logistic Function Estimation for CCPs
2018 Ohio Supreme Court Races

contest, appearing first on the ballot gives Donnolley a 3.13% advantage on average. In the race of DeGenaro versus Stewart, the advantage of being first is estimated to be 4.18% on average.

## 6.3 Abstaining and Attention Sets

Next, we take into account that many voters have not expressed a preference in these races even though they have done so in other races. Following our results in Section 4, we measure an identification region that allows for the fact that these voters could have had any ranking of the candidates. Finally, also following our previous results, we introduce attention sets by assuming that registered democrats and republicans only considered voting for a candidate of their own party. The following Figures illustrate the methods of Section 4 by combining instrumental variable, attention sets, and the presence of abstention. In Figures 10 and 11 the dark identification region uses as lower bound the fraction of voters who are registered for the candidate's party and as upper bound the conditional probability estimated using the regressions in Table 4; these regressions take into account the candidate's order on the ballot and therefore incorporate our instrumental variable methodology.

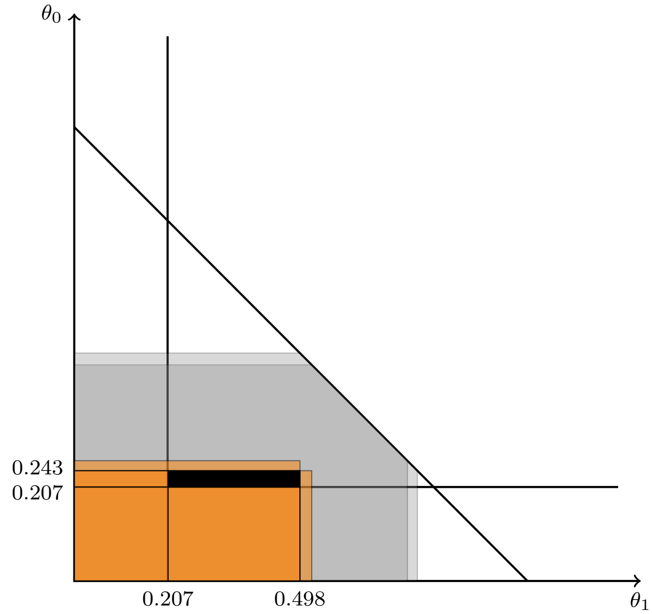



Figure 10: Identification Set: Baldwin vs. Donnelly

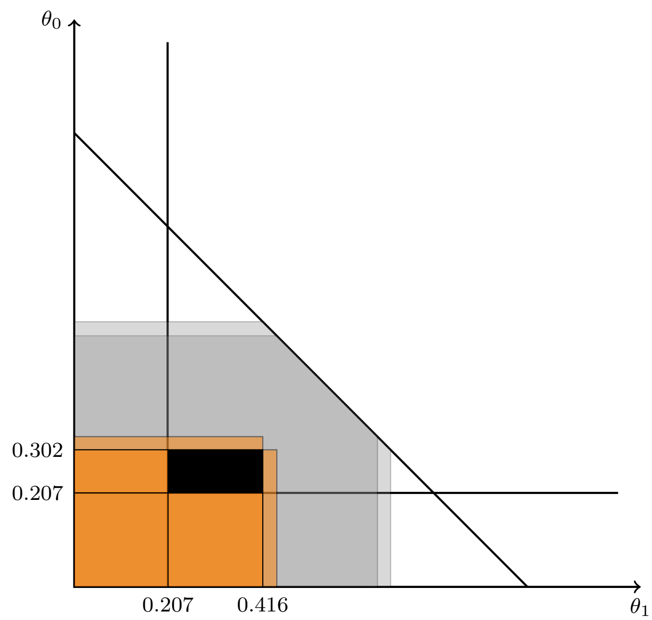



Figure 11: Identification Set: DeGenaro vs. Stewart

# 7 Conclusions

In this paper we provided the sharp identification region for a discrete choice model in which individuals' preference may not be complete and only aggregate choice data is available. The identification region is a strict subset of the parameter space, and thus disproves the idea that "everything goes when preferences are not complete". The identification region provides intuitive bounds on parameters of interest of the probability distribution of preferences across the population. Since our assumption do not rule out complete preferences, the identification region admits the two extreme possibilities of maximal incompleteness in which nobody can rank alternatives and no choice-relevant incompleteness in which everyone can rank alternatives. We illustrate how the existence of instrumental variables can rule out this last possibility.

Although we use interval orders as a way to describe preferences and behavior when incompleteness is allowed, our results extend beyond that model. In particular, theories in which behavior depends on two thresholds (instead of just one) can be accommodated into our framework. We have illustrated this possibility using the Knightian decision theory of Bewley (2002), but we believe our results would also extend to the multi-utility framework of Aumann (1962) and Dubra et al. (2004), or to the more recent twofold conservatism model of Echenique et al. (2021), as well as individual decision making models with "thick" indifference curves. All that one needs is to have a set of non-dominated alternatives that depends on two numbers per each alternative.

# Appendices

## A Random Sets

Identification analysis in economics has greatly benefited from tools developed in Random Set Theory (RST).[21] We use results developed in Beresteanu et al. (2011) and Beresteanu et al. (2012) to identify the quantities and parameters of interest. Specifically, we focus on the containment functional approach to partial identification.[22] We start with defining random sets and then the concepts related to the two identification strategies that are used in later sections of this paper.

The probability space, $(I, \mathcal{F}, \mathbf{P})$, on which all random variables and sets are defined is non-atomic. We use $i \in I$ to denote a random individual from the population $I$. From here on equalities and the statement "for every $i$" mean for every $i \in I$, $\mathbf{P}$-a.s. A random set is a measurable map defined as follows.

**Definition 4** *A random set $X$ is a mapping $X : I \to \mathcal{K}(\mathbb{R}^n)$ where $\mathcal{K}(\mathbb{R}^n)$ is the set of all closed subsets of $\mathbb{R}^n$ and such that for all $K \subset \mathcal{K}(\mathbb{R}^n)$ compact, $\{i : X(i) \cap K \neq \emptyset\} \in \mathcal{F}$.*

A random set can be thought of as a collection of point-valued random variables. The collection of all (point-valued) random variables, $x$, defined on our probability space such that $x(i) \in X(i)$ for all $i$ is defined as follows.

**Definition 5** *For a random set $X$, a selection of $X$ is a random variable $x$ such that $x(i) \in X(i)$ for all $i$. We let $Sel(X)$, the* selection set *of $X$, be the collection of all selections of $X$.*

### A.1 Containment Functional

The containment functional of a random set corresponds to the distribution function of a regular random variable.

**Definition 6** *For a random set $X$ and $\forall K \in \mathcal{K}(\mathbb{R}^n)$, the containment functional is defined as*

$$C_X(K) = \mathbf{P}(X \subset K).$$

[21]Molchanov (2005) presents a general exposition of RST. We focus on real-valued random variables and sets. The reader is referred to Appendix A of Beresteanu et al. (2012) and to Molchanov (2005) for a more in depth discussion of RST.

[22]Another approach is called the Aumann Expectation approach to partial identification. Beresteanu et al. (2012) discuss the merits of both approaches and make recommendations as to where each approach may have an advantage.

The following result, sometimes referred to as Artstein's Lemma,[23] establishes a relationship between the selection set and the containment functional.

**Theorem 5** *(Artstein's Inequalities) Let $X$ be a random set and let $Sel(X)$ be its selection set. Then $x \in Sel(X)$ if and only if*

$$C_X(K) \leq \mathbf{P}(x \in K)$$

*for all $K \in \mathcal{K}(\mathbb{R}^n)$.*

When a single selection $x$ from the random set $X$ is observed, $\mathbf{P}(x \in K)$ is identified from the data. It can then be used to draw restrictions on the possible values of the containment functional. This approach is especially useful when one looks at projections of the random set and its selections on a lower dimensional space as we show in section 3.

# B Data

The data used in Section 6 on 2018 midterm elections in Ohio was collected from the Ohio Board of Elections.[24] The data contains information on all the races including the positions, the names of the candidates, number of registered voters, votes cast for each candidates and the order in which the candidates appeared on the ballot. In cases where the party affiliation of the candidates was revealed on the ballot, we collected this information as well. Over all there were 8904 precincts in Ohio in the 2018 midterm elections. Some ballots included up to twenty different races depending on the district in which the precinct is located.

Midterm elections happen in the US every four years in between the General (presidential) elections. In general, midterm elections include three type of races; (1) National races (e.g. US senators and US congressmen), (2) State Races (e.g. governor and state supreme court judges) and (3) district/local races (e.g. state congress and board of education).

National races for positions like the governor of the state or the US senator are high profile races. The candidates in these races are affiliated with one of the two major parties - Democratic or Republican - and are well funded. The candidates use this money to widely advertise themselves and enjoy the support of their parties. As a result, voters are likely to be familiar with the names of these candidates when they come to vote. Possible exceptions are candidates who run for either the Green party or the the Libertarian party who are less widely known.

[23] See Molchanov (2005) Theorem 2.20 and Theorem 2.1 in Beresteanu et al. (2012).

[24] Most of the information is accessible from the Ohio Board of Elections web page at https://www.boe.ohio.gov. Additional information was obtained from county specific webpages, for example https://www.boe.ohio.gov/adams/election-info/ for Adams county.

On the ballot there are also races who receive less attention in the media. These positions, like auditor of the state, do not receive the same level of attention and campaign funding. As a result, the candidates running for these positions are less known. Candidates for these positions are affiliated with a party and their affiliation is denoted on the ballot. Ohio does not allow straight ticket party voting.

Finally, there are races where candidates do not have party affiliation. For example, state supreme court judges and state board of education. In these races, candidates tend to be both less familiar to the voter and cannot enjoy party affiliation or their party affiliation is not indicated on the ballots.

The order in which candidates appear on the ballot is as follows. Within each county the precincts are ordered by the precinct's code. In the first precinct candidates appear on the ballot by their alphabetical order. Then in the next precinct and on the candidate that appeared last in the previous precinct on moves to first place on the ballot and the other candidates move one spot down each.

# References

AUMANN, R. J. (1962): "Utility Theory without the Completeness Axiom," *Econometrica*, 30, 445–462.

BAYRAK, O. AND J. HEY (2020): "Understanding Preference Imprecision," *Journal of Economic Surveys*, 34, 154–174.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

——— (2012): "Partial Identification Using Random Set Theory," *Journal of Econometrics*, 166, 17–32.

BERRY, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262.

BEWLEY, T. F. (1986): "Knightian Decision Theory: Part I," Tech. rep., Cowles Foundation.

——— (2002): "Knightian Decision Theory: Part I," *Decisions in Economics and Finance*, 2, 79–110.

CALLANDER, S. AND C. H. WILSON (2006): "Context-dependent Voting," *Quarterly Journal of Political Science*, 1, 227–254.

DUBRA, J., F. MACCHERONI, AND E. A. OK (2004): "Expected Utility Theory without the Completeness Axiom," *Journal of Economic Theory*, 115, 118–133.

DZIEWULSKI, P. (2021): "A comprehensive revealed preference approach to approximate utility maximisation and non-transitive indifferences," Working Paper Series 0621, Department of Economics, University of Sussex Business Schoo.

ECHENIQUE, F., L. POMATTO, AND J. VINSON (2021): "Twofold Conservatism in Choice under Uncertainty," arxiv:2012.14557 [econ.th], arXiv.

ELIAZ, K. AND R. SPIEGLER (2011): "Consideration Sets and Competitive Marketing," *The Review of Economic Studies*, 78, 235–262.

FEDDERSEN, T. J. AND W. PESENDORFER (1999): "Abstention in Elections with Asymmetric Information and Diverse Preferences," *American Political Science Review*, 93, 381–398.

Fishburn, P. C. (1970): *Utility Theory For Decision Making*, Wiley and Sons.

Hall, M. G. and C. W. Bonneau (2008): "Mobilizing Interest: The Effects of Money on Citizen Participation in State Supreme Court Elections," *American Journal of Political Science*, 52, 457–470.

Hoderlein, S. (2011): "How many consumers are rational," *Journal of Econometrics*, 164, 294–309.

Horowitz, J. L. and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

——— (1998): "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58.

Kitamura, Y. and J. Stoye (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86, 1883–1909.

Krosnick, J. A., J. M. Miller, and M. P. Tichy (2004): "An unrecognized need for ballot reform: The effects of candidate name order on election outcomes," in *Rethinking the vote: The politics and prospects of American election reform*, ed. by A. Crigler, M. Just, and E. McCaffery, New York: Oxford University Press, chap. 4, 51–73.

Luce, R. D. (1956): "Semiorders and a Theory of Utility Discrimination," *Econometrica*, 24, 178–191.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer.

Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.

Manzini, P. and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

Marble, W. (2017): "Mail Voting Reduces Ballot Roll-Off: Evidence from Washington State," Tech. rep., Stanford, Department of Political Science.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *The American Economic Review*, 102, 2183–2205.

Matzkin, R. L. (1994): *Restrictions of Economic Theory in Nonparametric Methods*, Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 42.

——— (2007): *Nonparametric Identification*, Elsevier, vol. 6B of *Handbook of Econometrics*, chap. 73.

Meredith, M. and Y. Salant (2013): "On the Causes and Consequences of Ballot Order Effects," *Political Behavior*, 35, 175–197.

Miyashita, M. and Y. Nakamura (2021): "An Axiomatic Approach to Failures in Contingent Reasoning," Working paper series, Yale.

Molchanov, I. (2005): *Theory of Random Sets*, Springer.

Nevo, A. and A. Rosen (2012): "Identification with imperfect instruments," *Review of Economics and Statistics*, 93, 695–671.