

Identification of Incomplete Preferences

Arie Beresteanu*

Luca Rigotti†

January, 2020‡

Abstract

This 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. This connection is very natural because incomplete preferences models typically yield multiplicity of outcomes. In particular, we introduce an environment in which one can test the commonly held assumption that preferences satisfy the completeness axiom using data generated by the choices of heterogeneous individuals. We use methods from random set theory, and results from the literature on partial identification, to illustrate the connection between data and the distribution of utility values in the population.

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

Incomplete. Do not Cite. Do not distribute.

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

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

‡We thank Richard Van Weelden for pointers and encouragements. We thank Miriam Blumenthal for excellent RA work.

1 Introduction

This 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.¹ This connection is very natural because incomplete preferences models typically yield multiplicity of outcomes. In particular, we introduce an environment in which one can test the commonly held assumption that preferences satisfy the completeness axiom using data generated by the choices of heterogeneous individuals. One says preferences are complete when the decision maker can compare and rank any two alternatives at her disposal. Without completeness, decision makers may not be able to rank some alternatives. When this happens, behavior is indeterminate in the sense that any choice among these alternatives is possible. This indeterminacy is robust because incompleteness, unlike indifference, is not easy to avoid. This is because indifference between two alternatives is a knife-edge situation; it can be easily broken in the direction of either one. Incompleteness, however, is not as easy to avoid. Think, for example, of the incompleteness described by “thick indifference curves”; in that case, small changes to either alternative may not matter at all. For these reasons, indifference differs in a fundamental way from incompleteness, and one cannot disregard individuals who cannot compare alternatives. In bringing standard models to data, one typically argues that an individual who is indifferent among two alternatives is an extremely rare event because indifference is a knife-edge situation. Therefore, this possibility can be ruled out based on zero measure arguments. Without completeness, however, this reasoning no longer holds and we cannot disregard the possibility that many individuals are unable to rank the alternatives in front of them.

When using data to learn about the distribution of preferences in an heterogeneous population, allowing for incompleteness poses what may seem like an insurmountable problem. Since the theory says nothing about how choices are made between incomparable alternatives, data cannot distinguish between choices made because of a definite preference from choices made randomly because two alternatives are incomparable. While in standard discrete choice models this source of randomness is ruled out by the assumption that everyone’s preferences are complete, one may suspect that when completeness is not imposed data cannot say anything about preferences. To the contrary, we show that data can be used to learn about the distribution of preference parameters without assuming the preferences of all individuals are complete. The main difference is that instead of the usual point identification, one can only obtain partial identification of these parameters. We also show that, interestingly and surprisingly, under some conditions one may be able to rule out the hypothesis that all individuals in the population have complete preferences. These conditions are given by the existence of observables that are correlated with choices but orthogonal to preferences. When these instrumental variables exist, completeness is a testable assumption.

¹Incompleteness in decision making under risk and uncertainty was first studied by Aumann [1962] and Bewley [1986] (eventually published as Bewley [2002]), and extended by Dubra et al. [2004], Ok [2002]. Some applications of this work include Rigotti and Shannon [2005], and Lopomo et al. [2012].

In this paper we consider the simplest possible setting in which incompleteness can have observable consequences when looking at data collected by observing heterogeneous agents: a discrete choice model with two alternatives. In the standard model with complete preferences, each alternative is associated with a utility value which is known to the decision maker but not known to the analyst. One then uses observed choices to make inferences about the distribution of utility values in the population. This inference is driven by the assumption that choices are made in a utility maximizing fashion. Observing the proportion of individuals who choose one alternative over another tells us that the same proportion must assign higher utility to the corresponding alternative. In our initial model, we weaken this framework by allowing for the possibility that each alternative is associated with an interval of utility values. This corresponds to the *interval order* of Fishburn [1970], who calls the width of the interval associated to each alternative its *vagueness*. When the intervals are disjoint, the alternatives can be ranked and choice follows from utility maximization; when the intervals overlap, the corresponding alternatives are not comparable and choice is indeterminate.

We use methods from random set theory, and results from the literature on partial identification, to illustrate the connection between data and the distribution of utility values in the population. The observed proportion of individuals who choose one alternative implies bounds on the probability that any individual assigns a “higher” or “lower” interval to that alternative. Intuitively, these bounds follow from the observation that the fraction of individuals who chooses one alternative gives an upper bound to the fraction of individuals who assigns a higher interval to it. On the other hand, the data cannot impose a lower bound to this same fraction as it could be the case that all those who make one choice do it because they cannot compare it to the other possibility. Based on these results, we conclude that discrete choice data is informative about preference parameters even when incompleteness is allowed.

The bounds described above are, by their very nature, compatible with the possibility that all decision makers have complete preferences. As a matter of fact, one cannot rule out the possibility that everyone in the population has complete preferences. Intuitively, that is because individuals choose an alternative for two possible reasons that the data cannot disentangle: either that alternative has higher utility or it is not comparable to the other possible choices. As a natural response to this situation, we focus on finding conditions under which the data can actually rule out the possibility that everyone’s preferences are complete. In particular, we show that an instrumental variable approach can yield these conditions. If there is an observable variable that is orthogonal to preference parameters but correlated with choice, it can be used to rule out the possibility that all individuals have complete preferences.

Finally, we move to the most widely used model of incomplete preferences, the Knightian Decision Theory of Bewley [1986]. In the framework of decision making under uncertainty, that model derives a preference representation where alternatives are compared in an expected utility fashion, but using multiple probability distributions. One alternative is preferred to another if it has higher expected utility for all probability distributions in a closed

convex set. In this model, comparisons are done one probability distribution at a time; when all these comparisons agree, the two alternatives can be ranked. Unlike the models so far, Bewley’s theory does not rank alternatives using an interval of utility values for each. However, our previous results can be extended to this setting. Here, the data provides bounds on the probability that the highest and lowest expected utility difference between the two alternatives is negative.

Structure of the paper Section 2 reviews the two main ingredient of this paper: the theory of incomplete preferences and random set theory. Section 3 discusses random decision makers and define the identification set for parameters of preferences distributions. Section 4 covers binary choice models and Section 5 general discrete choice. Section 6 Extends these results and discuss additional assumptions that may lead to point identification in section 7 we apply our findings to a data set on voting in Ohio. Section 8 concludes.

2 Background

This paper combines two literatures. The first is the literature on partial identification using random set theory. The second is the literature on incomplete preferences. In this section we introduce the terminology and the basic results of these literatures that we use throughout this paper.

2.1 Random Sets

Identification analysis in economics has greatly benefited from tools developed in Random Set Theory (RST).² 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 two sets of tools for the identification analysis: the Aumann integral approach and the containment functional approach.³ In section 3.1 we use the containment functional whereas in section 3.2 the Aumann integral approach is easier to apply. 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.

Assumption 1 (probability space) *The probability space, (I, \mathcal{F}, 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 $i \in I$, P -a.s.

Definition 1 *A random set X is a mapping $X : I \rightarrow \mathcal{K}(\mathbb{R}^n)$ 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, $\{i : X(i) \cap K \neq \emptyset\} \in \mathcal{F}$.*

²Molchanov [2005] presents a general exposition of RST, while 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.

³Beresteanu et al. [2012] discuss the merits of both approaches and make recommendations as to where each approach may have an advantage. In this paper we use both approaches.

A random set can be thought of as a collection of point-valued random variables. For a random set X , we let $Sel(X)$, the *selection set* of X , be the collection of all (point-valued) random variables, x , defined on our probability space such that $x(i) \in X(i)$ for all i . Both identification approaches draw a relationship between a random set and its selections.

2.1.1 Containment Functional Approach

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

Definition 2 For a random set X and $\forall K \in \mathcal{K}(\mathbb{R}^n)$, the containment functional is defined as

$$C_X(K) = P(X \subset K).$$

The following theorem ties between the selection and the containment functional.

Theorem 1 (*Artstein Inequalities*) Let X be a random set and let $Sel(X)$ be its selection set. Then $x \in Sel(X)$ if and only if

$$C_X(K) \leq P_x(K)$$

for all $K \in \mathcal{K}(\mathbb{R}^n)$ where P_x is the probability distribution of the random variable x .

When a single selection x from the random set X is observed, $P_x(K)$ is identified from the data. It can then be used to draw restrictions on the values that the containment functional can take. This approach is especially useful in case we look at projections of the random set and its selections on a lower dimensional space. We use this approach in section 3.1.

2.1.2 Aumann Expectation Approach

The selection set leads to a natural definition of expectation of random sets.

Definition 3 Let X be a random set defined on a non-atomic probability space. If there is $x \in Sel(X)$ such that $E(x)$ is finite, then we say that the random set is integrable. The set of integrable selections of X is denoted as $Sel^1(X)$. The Aumann expectation of an integrable random set X is defined as

$$\mathbb{E}(X) = \{E(x) : x \in Sel^1(X)\}$$

The Aumann expectation of a random set produces a convex set (see Molchanov [2005] Theorem 2.1.15). Convex sets can be characterized by their support function.

Definition 4 Let $A \subset \mathbb{R}^n$ be a convex set and let \mathcal{B} be the unit sphere in \mathbb{R}^n . The support function $h : \mathcal{B} \rightarrow \mathbb{R}$ is defined as $h(u; A) = \sup_{a \in A} \langle u, a \rangle$ for all $u \in \mathcal{B}$.

The support function measures how far the set A stretches in each direction. For a selection x of X we know its expectation must lie within the Aumann expectation of the random set. Since the last is a convex set, we can verify the inclusion of $E(x)$ in $\mathbb{E}(X)$ using the support function. This approach is useful when the random set is characterized by a finite parameter. We use this approach in section 3.2.

2.2 Incomplete Preferences and Vagueness

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 alternative is driven by the utility value assigned to each. This follows from the well known result stating that, with a finite number of alternatives, for any complete and transitive preference relation there exist a utility function representing those preferences. This utility function assigns a value to each alternative, and these values can be used to describe choices made by a consumer who has those preferences. In other words, each alternative is associated with one number, and these numbers are used to model choices. Without completeness, this result no longer holds and one needs something that replaces it. For our purposes, any model in which each alternative is associated with two numbers, and these numbers determine choice behavior, will work. In what follows, we adopt a theory called interval orders (see Fishburn [1970]), but this is one of many possibilities.

We begin by presenting a model of decision making that allows preferences to be incomplete. Let \mathcal{A} be universal set of alternatives.⁴

Assumption 2 (partial order) *Let (\mathcal{A}, \succ) be a partially ordered set such that \mathcal{A} is a finite set and \succ is a strict preference relation over \mathcal{A} that satisfies the following two properties: (i) $\neg(a \prec a)$ (where \neg denotes logical negation), and (ii) $a \prec b$ and $c \prec d \Rightarrow a \prec d$ or $c \prec b$.*

The first property is irreflexivity, while the second is a strong form of transitivity.⁵ Fishburn [1970] calls such a preference an **interval order**, and it shows that \succ can be represented using two functions as the following theorem illustrates.

Theorem 2 (Fishburn (1970)) *If \succ is an interval order and \mathcal{A} is finite, then there exists two functions $\underline{u} : \mathcal{A} \rightarrow \mathbb{R}$ and $\sigma : \mathcal{A} \rightarrow \mathbb{R}$ with $\sigma(a) > 0$ for all $a \in \mathcal{A}$, such that*

$$a \prec b \quad \text{if and only if} \quad \underline{u}(a) + \sigma(a) < \underline{u}(b).$$

Thus, b is preferred to a if and only if the utility of b exceeds the utility of a 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,

⁴In a more general setting each decision maker has a set $A(i) \subset \mathcal{A}$ which is called the budget set and represents the alternatives feasible for decision maker i . Although our analysis does not rule out different budget sets for different agents, for simplicity we assume that all decision makers face the same set of alternatives.

⁵One can easily verify that (i) and (ii) immediately imply transitivity.

Fishburn calls σ a *vagueness* function since it measures the amount of imprecision associated with each alternative. The name interval order follows from the observation that alternatives are compared using intervals: when b is preferred to a the interval $[\underline{u}(b), \underline{u}(b) + \sigma(b)]$ lies to the right of the interval $[\underline{u}(a), \underline{u}(a) + \sigma(a)]$. Interval orders are clearly not necessarily complete. When the two intervals overlap, a and b are not comparable. In this setting, although strict preference is transitive, incomparability is not. In other words, a may be not comparable to b , and b maybe not comparable to c , but c is strictly preferred to a . This idea is sometimes referred to as intransitive indifference.⁶

In what follows, for ease of exposition, we let $\bar{u}(a) = \underline{u}(a) + \sigma(a)$ so that an 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.

When only two alternatives are present, the choice implied by an interval order is easy to describe. An alternative 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. With more than two alternatives, we extend this idea as follows: an alternative can be chosen from a particular list of options provided there is no other option in the list that is strictly preferred to it. We formalize this notion as follows.

Definition 5 *The subset of \mathcal{A} of **non-dominated alternatives** is defined as*

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

The set M is the set of all alternatives in \mathcal{A} that are not strictly dominated by another alternative. Without any further assumptions, the model may have multiple predictions. In other words, the model of discrete choice with interval utility is an incomplete model. We denote by y the observable choice (selection) made by the decision maker and make the following assumption.

Assumption 3 (Rationality) *Decision makers are rational and therefore, $y \in M$.*

With complete preferences, the set M includes alternatives the decision maker is indifferent among. Typical assumptions in empirical work ensure that M is a singleton and thus M and y are synonymous. With interval order the set M is not necessarily a singleton and may include several alternatives that the decision maker cannot rank. This captures the idea that, although choice can be indeterminate, only the undominated alternatives can be chosen by the individual. In particular, this reflects the idea that no selection rule is a-priori imposed among incomparable alternative, provided none of them is strictly worse. On the other hand, an alternative that is strictly dominated is never chosen.⁷

⁶Quoting 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 $x = 1000$."

⁷Using 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.

3 Random Decision Makers

The notation set in section 2 refers to a single decision maker. We now turn to a situation where there are many decision makers with heterogeneous preferences. We maintain the assumption that the universal set of alternatives faced by all decision makers is \mathcal{A} which satisfies assumption 2.

As in Section 2.1, let (I, \mathcal{F}, P) be a non-atomic probability space satisfying assumption (1). We use $i \in I$ to denote a random individual from the population I . Every decision maker $i \in I$ has an interval order over the set of choices \mathcal{A} that satisfies Assumption 2. By Theorem 2 we can say that for each $a \in \mathcal{A}$, decision maker i has a utility interval $[\underline{u}(a, i), \bar{u}(a, i)]$. We can, therefore, treat \underline{u} and \bar{u} as two random functions defined on \mathcal{A} such that for every $a \in \mathcal{A}$, $\Pr[\underline{u}(a) \leq \bar{u}(a)] = 1$. We are interested to learn about the joint distribution of utilities $\mathcal{U} = \{(\underline{u}(a), \bar{u}(a))\}_{a \in \mathcal{A}}$ or features of this distribution using the joint distribution of the observed variables. We denote the joint distribution of \mathcal{U} by $P_{\mathcal{U}}$.

For $i \in I$ we denote by $M(i)$ the set of alternatives that are not dominated.⁸ In other words,

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

Since \mathcal{A} is a finite set, $M(i)$ is non-empty and therefore M is a random set.

3.1 Nonparametric approach

Let θ denote a certain parameter of the joint distribution of \mathcal{U} . For example, $\theta = \Pr(\bar{u}(b) \leq \underline{u}(a))$ which is the probability mass of decision makers for which alternative a strictly dominates alternatives b . In this section we leave the distribution of \mathcal{U} unspecified. Let Θ be the set of all feasible values that θ can take. Θ is called the parameter space. In our example, $\Theta = [0, 1]$. In this section we show how can we use observational data to partially identify parameters of interest.

Consider the random set M defined in equation (1). The distribution of the random set M can be represented by its containment functional C_M . The values that the random set M takes and the value of its containment functional depend on the unknown joint distribution of \mathcal{U} .

Assumption 3, rationality, implies that the choice of individual i , y_i is an element of the set $M(i)$. In the context of random set theory, this behavioral assumption translates to the following measurability assumption.

Assumption 4 (measurability) *the random variable y is a selection of the random set M .*

Therefore, rationality implies that the random variable y is a selection from the random set M .

⁸The set $M(i)$ should be indexed by \mathcal{U} to denote that it depends on the joint distribution of \mathcal{U} which we want to identify. For simplicity we use $M(i)$ instead of $M_{\mathcal{U}}(i)$.

For any $K \subset \mathcal{A}$, the choice probabilities $\Pr(y \in K)$ are identified from the data generating process. Following Artstein's Lemma, we can say that y is a selection of M if and only if for all $K \subset \mathcal{A}$,

$$C_M(K) \leq \Pr(y \in K). \quad (2)$$

Computing the containment functional gives,

$$\begin{aligned} C_M(K) &= \Pr(M \subset K) \\ &= \Pr(M \cap K^c = \emptyset) \\ &= \Pr(\max_{k \in K^c} \bar{u}(k) < \max_{a \in \mathcal{A}} \underline{u}(a)) \\ &= \Pr(\max_{k \in K^c} \bar{u}(k) < \max_{a \in K} \underline{u}(a)). \end{aligned}$$

Inequality (2) imposes restrictions on the joint distribution of the random variables \mathcal{U} . Let \mathcal{P} be the family of all probability distributions over the set $\mathcal{U} = \{\underline{u}(a), \bar{u}(a)\}_{a \in \mathcal{A}}$ such that $\underline{u}(a) \leq \bar{u}(a)$ for all $a \in \mathcal{A}$ with probability

1. Given our knowledge of $\{\Pr(y \in K)\}_{K \subset \mathcal{A}}$, the identification region is defined as

$$\mathcal{P}^I = \{P \in \mathcal{P} : C_M(K) \leq \Pr(y \in K) \forall K \subset \mathcal{A}\}. \quad (3)$$

Artstein's Lemma implies that the set \mathcal{P}^I is the sharp identification region for the set of joint distributions on \mathcal{U} . Appendix A provides two results that shows how essential are the inequalities implied by Artstein's Lemma.

With no further restriction on the joint distribution of \mathcal{U} beyond $\underline{u}(a) \leq \bar{u}(a)$, we cannot say anything further. There are two possible ways to make progress. First, the joint distribution of \mathcal{U} can be assumed to belong to a finite parameter family and we can seek to partially identify this parameter. We take this approach in Section 3.2. Second, we can avoid parametric assumptions but limit our interest to a finite number of features of the joint distribution of \mathcal{U} . We proceed in this direction in Sections 4 and 5.

3.2 Parametric approach

Suppose first that the joint distribution of \mathcal{U} is characterized by a finite dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. We denote this distribution as P_θ .

For $\theta \in \Theta$, let $\underline{u}_\theta(a, i)$ and $\bar{u}_\theta(a, i)$ be the lower and upper utilities for decision maker i for alternative $a \in \mathcal{A}$ given the parameter is θ . For $i \in I$ we denote by $M_\theta(i)$ the set of alternatives that are not dominated. In other words,

$$\begin{aligned} M_\theta(i) &= \{a \in \mathcal{A} : \nexists b \in \mathcal{A} \text{ such that } \underline{u}_\theta(b, i) > \bar{u}_\theta(a, i)\} \\ &= \{a \in \mathcal{A} : \max_{b \in \mathcal{A}} \underline{u}_\theta(b, i) < \bar{u}_\theta(a, i)\}. \end{aligned} \quad (4)$$

The identification region in equation (3) can be written as

$$\Theta^I = \{\theta \in \Theta : C_{M_\theta}(K) \leq P_\theta(y \in K) \forall K \subset \mathcal{A}\}. \quad (5)$$

For each $m \in \text{Sel}(M)$, we define a random variable $q(m)$ who gets its values in $\Delta^{|\mathcal{A}|}$ such that $[q(m(i))]_k = 1[m(i) = k]_{k=1..|\mathcal{A}|}$ and $1[\cdot]$ is the indicator function. In other words, $q(m)$ is a random vector whose entries all equal to zero except from the coordinate that represents the choice made by i . The expected value of this random vector $E(q(m))$ is a point in the simplex $\Delta^{|\mathcal{A}|}$ represents the distribution of choices.

We define the following random set

$$Q(M) = \{q(m) : m \in \text{Sel}(M)\}. \quad (6)$$

The random set $Q(M)$ represents all the possible choices that our incomplete model can predict. $Q(M)$ is a random set who takes its values in $\Delta^{|\mathcal{A}|}$. Hence, the Aumann integral of $Q(M)$ is well defined,

$$\mathbb{E}(Q(M)) = \{E(q) : q \in \text{Sel}(M)\}. \quad (7)$$

We say that the set $\mathbb{E}(Q(M))$ is the set of model predictions for choice distributions. This set includes all the distributions over the finite set \mathcal{A} that may result from the distribution of incomplete preferences \mathcal{U} .

Let y_i denote the observed choice made by i , such that $y_i \in \mathcal{A}$. We may also observe additional vectors of relevant covariates for individual i depending on the context. For simplicity we avoid conditioning on these variables in this section. The observed choice, y is assumed to satisfy Assumption 3, rationality. In other words, we assume that $y \in \text{Sel}(M)$. Using Theorem 2.1 in BMM (2012), the identification region for $P_{\mathcal{U}}$ is,

$$\Theta^I = \{P : \min_{v \in \mathcal{B}} (E[h(Q(M), u)] - u'P(y))\} \quad (8)$$

where \mathcal{B} is the unit ball in $\mathbb{R}^{|\mathcal{A}|-1}$ and $h(A, v)$ is the support function of a set $A \subset \mathbb{R}^{|\mathcal{A}|-1}$ evaluated in direction $v \in \mathcal{B}$.

4 Binary Choice

In this section we focus on a binary choice models. Let the set of alternatives be $\mathcal{A} = \{a_0, a_1\}$ for all i . Let $u_i(a) = [\underline{u}(a), \bar{u}(a)]$. This is the interval utility that agent i assigns to alternative $a \in \mathcal{A}$. This model assumes that the interval is known to the decision maker and thus she can maximize her utility.

Let $y_i = \mathbf{1}[u_i(a_1) > u_i(a_0)]$. We denote the choice probability of alternative a_1 by $p_1 = \Pr(y_i = a_1)$ and the choice probability of a_0 by $p_0 = \Pr(y_i = a_0)$. It is clear that $p_1 = \Pr(V > 0)$ and $p_0 = \Pr(V < 0)$.⁹ Typically y is observed while V is known only to the agent but is unobserved by us. Therefore, the DGP identifies p_0 and p_1 and thus it identifies $\Pr(V < 0)$ and $\Pr(V > 0)$ respectively. In the literature on discrete choice it is common to specify a distribution function for V which is known up to a finite parameter θ . In this case, θ can be consistently and efficiently estimated by maximum likelihood.

⁹We assume that V is a continuous variable and thus $\Pr(V = 0) = 0$ and thus we can ignore this event or add it to any other event.

Suppose now that $V = [V_L, V_U]$ where V_L and V_U are two random variables such that $\Pr(V_L \leq V_U) = 1$. We fix $u(0)$ to be a singleton. In practice the decision maker can face vagueness with respect to the utility from choice 0 as well. Since what matters is the order of the utilities resulting from alternatives 0 and 1, we can think about V as the Hukuhara difference (see Hukuhara [1967]) between the random sets representing the utility from alternative 1 and alternative 0.

Given this vagueness about the utility from alternative 1, the decision maker's choice set is

$$Y = \begin{cases} \{1\} & V_L > 0 \\ \{0, 1\} & 0 \in [V_L, V_U] \\ \{0\} & V_U < 0 \end{cases}.$$

When $0 \in [V_L, V_U]$ the decision maker is unsure how to rank the two options she faces. Figure 1 illustrates the decision rule.

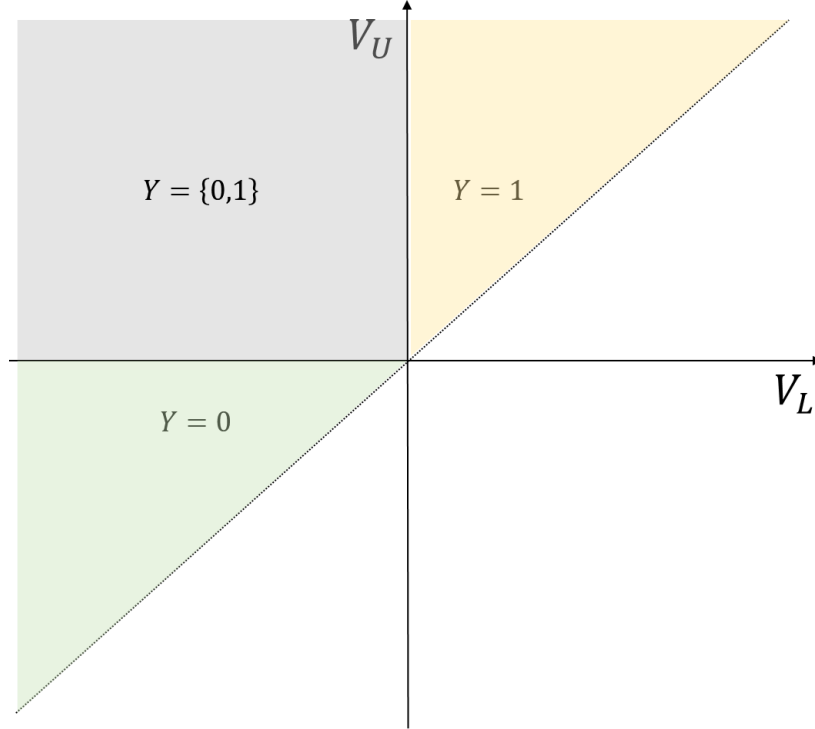


Figure 1: Choice Rule with Incomplete Preferences

As above, we denote the agent's choice by y . Following Artstein inequality ¹⁰ $y \in \text{Sel}(Y)$ if and only if $\Pr(y \in K) \geq C_Y(K)$ for every closed set K where $C_Y(K)$ is the containment functional defined as $C_Y(K) = \Pr(Y \subset K)$. Since $C_Y(\{0\}) = \Pr(V_U < 0)$ and $C_Y(\{1\}) = \Pr(V_L > 0)$, for any $y \in \text{Sel}(Y)$, $\Pr(y = 0) \geq \Pr(V_U < 0)$ and $\Pr(y = 1) \geq \Pr(V_L > 0)$. Using the fact that $\Pr(y = 0) = 1 - \Pr(y = 1)$, we can conclude that

$$\Pr(V_U < 0) \leq \Pr(y = 0) \leq \Pr(V_L < 0). \quad (9)$$

¹⁰See Molchanov [2005] Theorem 2.20 and Theorem 2.1 in Beresteanu et al. [2012].

Without observing y the probability $\Pr(y = 0)$ is unidentified and inequality (9) implies only that $\Pr(V_U < 0) \leq \Pr(V_L < 0)$. This is immediate from our initial assumption that $\Pr(V_L \leq V_U) = 1$. Suppose we observe data on choices made by decision makers facing the type of incompleteness described above. Since we can identify p_1 and p_0 from the data, inequality (9) becomes

$$\Pr(V_U < 0) \leq p_0 \leq \Pr(V_L < 0). \quad (10)$$

Without further assumptions this is the only information we can draw about the joint distribution of V_L and V_U . We cannot reject, for example, the hypothesis that $V_L = V_U$, $P - a.s.$ and there is no vagueness about choice 1. Moreover, the above inequality bound the marginal cumulative distributions of V_L and V_U at the point 0 but does not provide information on their joint distribution or other features of their marginal distributions. Another parameter of interest here is the proportion of decision makers who have incomplete preferences. This proportion is $\Pr(V_L < 0 \wedge V_U > 0) = 1 - \Pr(V_L > 0) - \Pr(V_U < 0) = \Pr(V_L < 0) - \Pr(V_U < 0)$. Again, this quantity can be anywhere between zero and one.

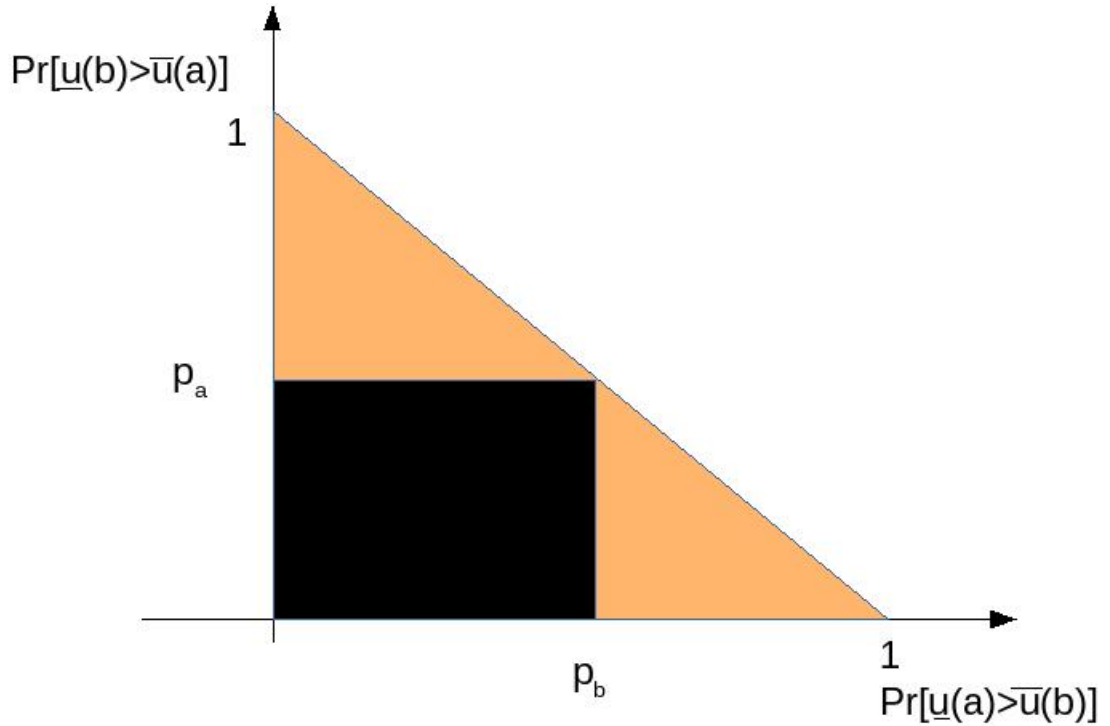


Figure 2: Partial identification with Two Alternatives

Figure 2 illustrates the information contained in inequality (10). Before observing any data the only information we have is that $V_L \leq V_U$ for every individual i . Therefore, $\Pr(V_L < 0) \geq \Pr(V_U < 0)$. The area under the 45° line represents this information. After observing data we can identify the frequency $p_0 = \Pr(y = 0)$. Inequality (10) then puts restrictions on the values that the cumulative distributions of V_L and V_U can take

at the point 0. For example, say $p_0 = 0.37$. We can no longer admit the case where $\Pr(V_L < 0) = 0.3$ and $\Pr(V_U < 0) = 0.25$. Similarly we cannot admit the case where $\Pr(V_L < 0) = 0.5$ and $\Pr(V_U < 0) = 0.4$. The two triangles corresponding to these cases are eliminated from the identification region after observing data. The remaining identification region is the rectangle north-west of the point (p_0, p_0) . Note that the case where $\Pr(V_L < 0) = \Pr(V_U < 0) = 0.37$ is included in the identification region. The fact that these two probabilities are equal to each other does not necessarily mean that there is no incompleteness and that $V_L = V_U$ for all i . It is still possible that the quality of choice 1 is not known to the agent precisely but $\Pr(0 < V_L < V_U) = 0.37$ and $\Pr(V_L < V_U < 0) = 0.63$. In other words, there is incompleteness but it is choice irrelevant.

The point (p_0, p_0) in Figure 2 represents a situation where $\Pr(V_U < 0) = p_0 = \Pr(V_L < 0)$ and therefore no DM has incomplete preferences. On the otherhand the point $(1, 0)$ represents a situation where $0 = \Pr(V_U < 0) < \Pr(V_L < 0) = 1$ and all DMs cannot rank the two alternatives. It is natural to ask what is the identification region if we know that at least a proportion $\alpha \in (0, 1)$ of decision makers cannot rank the alternatives. Using the same reasoning as above, we have $\alpha \leq \Pr(V_L < 0 \wedge V_U > 0) = 1 - \Pr(V_L > 0) - \Pr(V_U < 0) = \Pr(V_L < 0) - \Pr(V_U < 0)$. Therefore, this information adds the restriction that

$$\Pr(V_U < 0) \leq \Pr(V_L < 0) - \alpha. \quad (11)$$

Figure ?? depicts the identification region consistent with both inequalities (10) and (11). The dark region in Figure ?? includes all the points which are consistent with the statement that at least an α proportion of the DMs cannot rank the two alternatives in addition to the observation that p_0 portion of DMs eventually chose alternative 0.

The theory we have presented so far makes few assumptions. Our results illustrate that identification, albeit partial, is possible when preferences are not complete. Next, we illustrate how further assumptions can lead to even stronger conclusions in terms of identification. First, we briefly discuss assumptions on the joint distribution of (V_L, V_U) . Then, we show that the presence of particular instruments can lead to a rejection of the hypothesis that all consumers' preferences are complete. Finally, we discuss a possible way to break incompleteness using the idea of regret.

Example 1 Let $A = \{0, 1\}$. Let $\underline{u}(0) = \bar{u}(0) = 0$. Let $\underline{u}(1) = \beta - \varepsilon$ where $\varepsilon \sim F_\varepsilon$ such that F_ε is known and $E[\varepsilon] = 0$. Let $\bar{u}(1) = \underline{u}(1) + \sigma$ where $\sigma \in \mathbb{R}_+$. Let $y \in A$ be the choice made under rationality. We observe $p_0 = \Pr(y = 0)$ and $p_1 = \Pr(y = 1)$. Let H be the set of all pairs (β, σ) that are consistent with the observed (p_0, p_1) .

Case 1 (no vagueness): Assume $\sigma = 0$. $p_1 = \Pr(y = 1) = \Pr(\beta - \varepsilon > 0) = \Pr(\varepsilon < \beta) = F_\varepsilon(\beta)$. Therefore, $\beta = F_\varepsilon^{-1}(p_1)$. The identification set is a singleton

$$H = \{(F_\varepsilon^{-1}(p_1), 0)\}.$$

Case 2 (vagueness): $\sigma \in \mathbb{R}_+$. We know that $p_1 \geq \Pr(y(1) > 0) = \Pr(\beta - \varepsilon > 0) = \Pr(\varepsilon < \beta)$. We also know that $p_0 \geq \Pr(\bar{u}(1) < 0) = \Pr(\beta + \sigma - \varepsilon < 0) = \Pr(\varepsilon > \beta + \sigma) = 1 - F_\varepsilon(\beta + \sigma)$. These inequalities imply that $F_\varepsilon(\beta) \leq p_1 \leq 1 - F_\varepsilon(\beta + \sigma)$. In other words, $\beta \leq F_\varepsilon^{-1}(p_1) \leq \beta + \sigma$. The identification set is

$$H = \{(\beta, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \beta < F_\varepsilon^{-1}(p_1) < \beta + \sigma\}.$$

This identification region contains the point $(F_\varepsilon^{-1}(p_1), 0)$.

Case 3 (instrumental variable): As in Case 2, $\sigma \in \mathbb{R}_+$ but now we observe a random variable Z with support Z such that Z does not affect the valuation but it may affect the choice in case of incompleteness. In other words, β is not a function of Z and ε is independent of Z . We observe $p_{0|z} = \Pr(y = 0|Z = z)$ and $p_{1|z} = \Pr(y = 1|Z = z)$ for all $z \in Z$. Following the same analysis as above we can conclude that $\beta \leq \min_{z \in Z} F_\varepsilon(p_{1|z})$ and $\max_{z \in Z} F_\varepsilon^{-1}(p_{1|z}) \leq \beta + \sigma$. The identification set is

$$H = \{(\beta, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \beta \leq \min_{z \in Z} F_\varepsilon(p_{1|z}), \max_{z \in Z} F_\varepsilon^{-1}(p_{1|z}) \leq \beta + \sigma\}.$$

We can immediately see that if $\max_{z \in Z} F_\varepsilon^{-1}(p_{1|z}) > \min_{z \in Z} F_\varepsilon(p_{1|z})$, we can rule out all points for which $\sigma = 0$. In other words, if for different values of Z there are different choice probabilities this can only be a result of ambiguity.

Example 2 Let $A = \{a_0, a_1, a_2\}$ be the alternatives set. Suppose we normalize $u(a_0) = \bar{u}(a_0) = 0$. This is not a trivial normalization but we make this assumption for two reasons. First, we need a location and scale normalizations. This can be achieved by fixing the utility of alternative a_0 to zero and by setting the error terms distribution to one that has a fixed variance. Second, to have a tractable example, we want to reduce the number of parameters to be evaluated. For the other alternatives in A we assume

$$u(a_j) = \beta_0 + \beta_1 X_j - \varepsilon_j, \bar{u}(a_j) = u(a_j) + \delta_j,$$

where $j = 1, 2$. The parameter to be evaluated is $\theta = \{\beta_0, \beta_1, \delta_1, \delta_2\} \in \mathbb{R}^2 \times \mathbb{R}_+^2$. For simplicity we assumed that the intercept and slope in V_L are the same for both alternatives a_1 and a_2 . To make this example tractable we assume that ε_1 and ε_2 are independent of each other and independent of the covariate X . If we assume that $(\varepsilon_1, \varepsilon_2)$ are jointly Normal with vector mean 0 and the identity variance-covariance matrix $I_{2 \times 2}$, we have a Probit model with quality ambiguity. A similar extension of the Logit model can be made if we assume a T1EV distribution instead. Denote the cumulative distribution of ε_j as Φ for both $j = 1, 2$. To simplify notation, we are omitting conditioning on (X_1, X_2) in all probabilities below. Similarly to the discussion above we can see that

$$\begin{aligned} 1 - \Pr(\bar{u}(a_j) \geq \max_{k \in \{j\}^c} \bar{u}(a_k)) \\ &\geq p_j = \Pr(y = j) \\ &\geq \Pr(u(a_j) \geq \max_{k \in \{j\}^c} u(a_k)). \end{aligned}$$

Applying our assumptions we get

$$\begin{aligned} & \Phi(\beta_0 + \beta_1 X_1 + \delta_1) \cdot \Phi(\beta_0 + \beta_1 X_2 + \delta_2) \\ & \leq p_0 \leq \\ & 1 - [1 - \Phi(\beta_0 + \beta_1 X_1)] \cdot [\Phi(\beta_0 + \beta_1 X_2)] \end{aligned}$$

and similarly we can find inequalities for the other probabilities.

Numerical example

Consider the following model. The alternatives set is $A = \{a_0, a_1, a_2\}$. We make the following normalizations. For $i \in \{0, 1, 2\}$,

$$\mathbf{u}(i) = \beta_i + \varepsilon_i \tag{12}$$

$$\bar{u}(i) = V_L(i) + \sigma + \eta_i \tag{13}$$

We set $\beta_0 = 0$ and $\sigma = 1$. Therefore, we have only two parameters to identify: β_1 and β_2 . We made the following parameter choice,

$$\beta = (0, 0.094, -0.054) \tag{14}$$

We let ε_i and η_i be i.i.d $U[-0.25, 0.25]$. A data of 7250 observations simulated from the above model was generated. In other words, for each of the simulated 7250 individuals we generated 3 intervals represented their utility intervals for the 3 alternatives. We then had to make an arbitrary choice how these decision makers break ambiguity. There are of course many ways to do that. Our choice was to compute the mid-point of the 3 intervals and to pick the alternative with the highest value. As a result the sample we generated produced the following choice probabilities vector for the 3 alternatives,

$$p = (0.181, 0.766, 0.053). \tag{15}$$

This vector is then passed to the next stage where we identify which pairs of To check whether a certain pair of parameters is in the identification region we have to verify 6 inequalities. To do that we calculate the capacity functionals by simulation and compare them to the vector of choice probabilities above. The following scatter plot was generated using a differential evolution algorithm.

5 Discrete Choice with Incomplete Preferences

Our leading model is discrete choice. Denote the set of alternatives that each decision maker faces as $\mathcal{A} = \{a_0, a_1, \dots, a_J\}$. Let $V : I \rightarrow \mathbb{R}^{\mathcal{A}}$ denote the (random) utility vector that each decision maker i assigns to the alternatives in \mathcal{A} . For simplicity, we denote the vector of utilities for i as $V_i = (V_{i0}, V_{i1}, \dots, V_{iJ})$. We denote the

choice of decision maker i by $y_i = \arg \max_{j \in \mathcal{A}} V_{ij}$. We also denote the choice probabilities by $p_j = \Pr(y_i = j)$. If all V_{ij} are single valued, this is the regular discrete choice model where $p_j = \Pr(V_{ij} > V_{ij'} \forall j' \neq j)$.¹¹

The assumption that V_{ij} is single valued corresponds to the assumption that preferences are complete and any pair of alternatives is comparable for each decision maker. Instead, we lift the completeness assumption and assume that V_{ij} are potentially interval valued. The *interval order* axiomatized by Fishburn [1970] that is introduced in Section 2 is a notable example for interval valued utility.¹² In this model, $V_{ij} = [u_i(j), u_i(j) + \sigma(j)]$ and the function $u(\cdot)$ represents the lower bound of the utility interval. If for two different alternatives, j and j' , the corresponding utility intervals overlap, j and j' are not comparable. Fishburn calls σ a *vagueness* function since it measures the imprecision in a preference order.

5.1 Assumptions on the joint distribution of utility intervals

Suppose further that $(V_L, V_U) \sim F(\cdot, \cdot; \theta)$ where $\theta \in \Theta$ a finite dimensional parameter space. Under this assumption, the identification region is

$$\Theta_I = \{\theta \in \Theta : F_L(0; \theta) \leq p_0 \leq F_U(0; \theta)\} \quad (16)$$

where $F_L(t; \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^t F(u, v; \theta) dv du$ and $F_U(t; \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^t F(u, v; \theta) du dv$ are the marginal distributions of V_L and V_U respectively. Note that parts of θ that do not appear in either F_L or F_U cannot be identified at all.

Models of binary choice often make the assumption that $u_{1i} = V + \varepsilon$ where V is the average utility of choice 1 and ε has a parametric distribution F_ε with mean zero and a known constant variance. Two prominent choices for F_ε are the standard normal distribution leading to a Probit model and the logistic distribution leading to the Logit model. In both these cases, there are no unknown parameters in F_ε and the only unknown parameter is V . Extending this model to accommodate incompleteness can be done in several ways. Here we assume that the decision maker is unsure about the average utility of choice 1. Specifically we assume that $V_L = \bar{V}_L + \varepsilon$ and $V_U = \bar{V}_U + \varepsilon$. Under this model the identification region for (V_L, V_U) is

$$\Theta_I = \{(\bar{V}_L, \bar{V}_U) : \bar{V}_L \leq F_\varepsilon^{-1}(p_0) \leq \bar{V}_U\}. \quad (17)$$

Inequality (17) imposes restrictions on the possible values of the pair (\bar{V}_L, \bar{V}_U) . Again, the weak inequalities in (17) imply that the hypothesis $\bar{V}_L = \bar{V}_U$ cannot be rejected even in this simple model. Moreover, the difference $\bar{V}_U - \bar{V}_L$ is unbounded. Note, however, that inequality (17) is written in terms of the values (\bar{V}_L, \bar{V}_U) and not in terms of their distribution functions as in inequality (10). If ε is symmetric around 0, then if $p_0 < \frac{1}{2}$, $\bar{V}_L < 0$. If on the other hand $p_0 > \frac{1}{2}$, $\bar{V}_U > 0$. In other words, this model allows us to sign one value of the pair (\bar{V}_L, \bar{V}_U) . Furthermore, if we observe a vector of covariates X and say that (\bar{V}_L, \bar{V}_U) are each a linear function of the

¹¹We assume that V_{ij} are continuously distributed and thus the probability of ties is zero.

¹²A strict preference \prec is an interval order if it is irreflexive (not $x \prec x$) and satisfies a weak form of transitivity ($x \prec y$ and $z \prec w \Rightarrow x \prec w$ or $z \prec y$).

covariates, we can create an interval Logit or Probit. Further developing this model is part of our research agenda.

5.2 Instrumental variables

5.2.1 Perfect Instruments

In this section we show how one could use instrumental variables to further refine the identification region established in the general case. In particular, valid instruments have the characteristic of influencing choice between incomparable alternatives while having no effect on preferences. One example of these can be thought of as “frames” in the sense of behavioral economics. This is the well documented phenomenon that sometimes presenting the same information in differing fashion may lead to different choices. If these instruments exist, they can be used to rule out the possibility that all decision makers have complete preferences.

Suppose there is a variable Z with a support \mathcal{Z} such that (V_L, V_U) are independent of Z . The choice y , however, may depend on the instrument. For every $z \in \mathcal{Z}$, let $p_{0|z} = \Pr(y = 0|Z = z)$. This means that when the decision maker cannot compare the two alternatives, the choice between 0 and 1 may depend on Z .

Theorem 3 *Suppose (V_L, V_U) is independent of Z . Then,*

$$\Pr(V_U < 0) \leq \inf_{z \in \mathcal{Z}} p_{0|z} \wedge \sup_{z \in \mathcal{Z}} p_{0|z} \leq \Pr(V_L < 0).$$

If Z affects the choice, then $\inf_{z \in \mathcal{Z}} p_{0|z} < \sup_{z \in \mathcal{Z}} p_{0|z}$. In this case we can reject the hypothesis that $V_L = V_U$ and confirm that some decision makers have incomplete preferences. Figure 3 shows that the identification region for $(\Pr(V_U < 0), \Pr(V_L < 0))$ in the presence of an instrumental which is the dark shaded area does not touch the 45° line. Therefore, the amount of vagueness can be measured by $\Pr(V_L < 0) - \Pr(V_U) \geq \sup_{z \in \mathcal{Z}} p_{0|z} - \inf_{z \in \mathcal{Z}} p_{0|z}$.

Moreover, if we assume that there is a vector of covariates $x \in \mathbb{R}^k$ such that $V_L = x'\gamma_L + \varepsilon$ and $V_U = x'\gamma_U + \varepsilon$. In this case the identification region is

$$\Theta_I = \left\{ (\gamma_L, \gamma_U) : x'\gamma_U \leq \inf_{z \in \mathcal{Z}} F_\varepsilon^{-1}(p_{0|z}) \wedge \sup_{z \in \mathcal{Z}} F_\varepsilon^{-1}(p_{0|z}) \leq x'\gamma_L \right\}$$

5.2.2 Imperfect Instruments

In this section we relax the strong assumption that (V_L, V_U) is independent of Z . Nevo and Rosen [2012] introduced the notion of imperfect instrumental variables in a linear regression model with endogenous regressors. An imperfect instrumental variable is a situation where a variable Z is correlated with the error term of the regression but to a much lesser degree than its correlation with the endogenous regressor. Nevo and Rosen [2012] show that under some conditions this situation lends itself to partial identification of the regression parameters. We adjust this notion of imperfect instrumental variables to our model of choice with vagueness. We define the

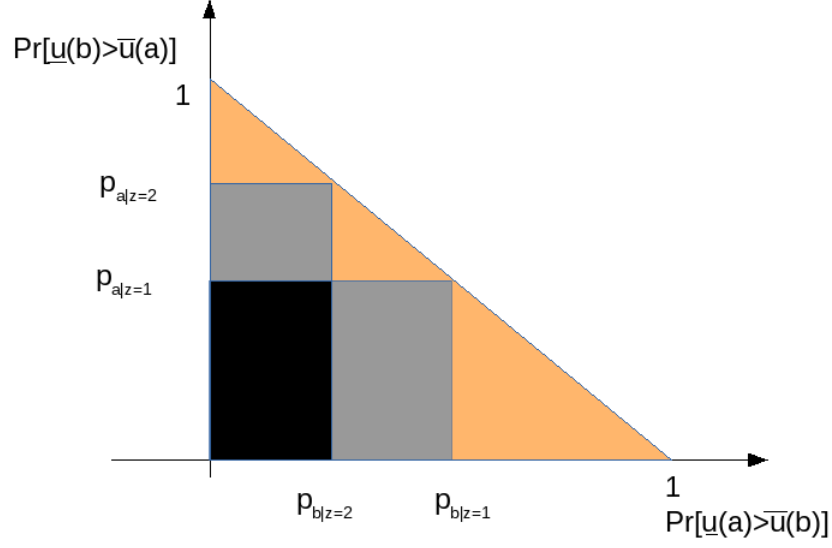


Figure 3: ID region with an Instrumental Variable

following quantities.

$$\begin{aligned}\delta_L &= \sup_{z \in \mathcal{Z}} |\Pr(V_L < 0|Z = z) - \Pr(V_L < 0)| \\ \delta_U &= \sup_{z \in \mathcal{Z}} |\Pr(V_U < 0|Z = z) - \Pr(V_U < 0)|\end{aligned}$$

and

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

Theorem 4 Suppose $\Delta_0 > \delta_L + \delta_U$, then $\Pr(V_U < 0)$ is bounded away from $\Pr(V_L < 0)$ by a strictly positive quantity.

Proof. For all $z \in \mathcal{Z}$,

$$\Pr(V_U < 0) - \delta_U \leq \Pr(V_U < 0|Z = z) \leq p_{0|z}$$

and thus,

$$\Pr(V_U < 0) \leq p_{0|z} + \delta_U.$$

As a result we can say that

$$\Pr(V_U < 0) \leq \inf_{z \in \mathcal{Z}} p_{0|z} + \delta_U.$$

Similarly, since

$$p_{0|z} \leq \Pr(V_L < 0|Z = z) \leq \Pr(V_L < 0) + \delta_L.$$

Using the same steps as above we get that

$$\sup_{z \in \mathcal{Z}} p_{0|z} - \delta_L \leq \Pr(V_L < 0).$$

It is now clear that

$$\Pr(V_U < 0) - \Pr(V_L < 0) > \Delta_0 - \delta_L - \delta_U > 0.$$

■

5.3 Minmax regret

The observed choice, y , is a selection from the random set Y . We could establish only very general bounds on the distribution of (V_L, V_U) because we made no assumptions on the selection y in case in which the consumer cannot compare alternatives. So far, we have been silent about the way the decision maker makes her choice when $Y(\omega) = \{0, 1\}$. This approach resulted in partial identification of the parameters of interest as described above. Next, we focus on the idea of minimizing maximal regret as a way to break the consumer's indecision.

Suppose that $(V_L, V_U) \sim F$ for some joint distribution F (it can be a parametric distribution as before). If $V_U < 0$, the decision maker chooses choice 0. Similarly if $V_L > 0$, the decision maker chooses choice 1. The interesting case occur when $V_L < 0 < V_U$. Both choices in this case may be rational. To break the tie the decision maker can use the minimax regret rule. In other words, we assume that y_i is such that i minimizes the maximum regret she can have when the true value of choice 1 is realized. If i chooses 1, the maximal regret is $|V_L|$ and if i chooses 0, the maximal regret is V_U . Therefore, i chooses 1 if $V_U \geq |V_L|$ and chooses 0 otherwise. This is summarized in the following equation. For each i ,

$$y = \begin{cases} 1 & \text{if } V_L > 0 \text{ or } (V_L < 0 < V_U \text{ and } V_U > |V_L|) \\ 0 & \text{if } V_U < 0 \text{ or } (V_L < 0 < V_U \text{ and } V_U < |V_L|) \end{cases} \quad (18)$$

In other words, the minimax regret rule picks one selection from the random set Y . Therefore, $p_0 = \Pr(V_U < 0 \text{ or } (V_L < 0 < V_U \text{ and } V_U < |V_L|))$. Figure ?? describes the choice rule in (18). As one can see, the region where choice was indeterminate, the ortant where V_U is positive and V_L is negative, is now allocated to each alternative so that points above the 45 degree line mean alternative 1 is chosen and point below it mean alternative 0 is chosen. Adding a parametric assumption on the joint distribution of (V_L, V_U) will allow us to point identify the finite parameter of that distribution. Specifically if we assume that $V_L = \bar{V}_L + \varepsilon$ and $V_U = \bar{V}_U + \varepsilon$ and $\varepsilon \sim F_\varepsilon$ for a known distribution F_ε , then we can identify the pair (\bar{V}_L, \bar{V}_U) .

6 Discrete Choice With Knightian Uncertainty

Next, we discuss how this framework can be extended to allow for ambiguity and Knightian uncertainty as described in Bewley [1986].¹³ That paper 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.¹⁴ Lack

¹³Bewley's original paper has been published recently as Bewley [2002].

¹⁴Incompleteness in decision making in a Von-Neumann and Morgenstern environment was first studied by Aumann [1962], which obtains a representation in which the probability is unique but there are multiple utility functions.

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, and when the preference relation is complete this set becomes a singleton.

We let S denote the state space and, with abuse of notation, also the cardinality of that space. There are J possible choices, and the utility of each choice depends on the state that is realized. Therefore, for each choice $j \in J$ we specify the utility that choice yields in each state. Formally, the utility of an individual for choice j in state s is denoted as $u(j, s)$. Let $\Delta(S)$ be the set of all probability distributions over S . If $\pi \in \Delta(S)$, the expected utility of individual i according to that probability is given by

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

Bewley's theory provides axioms on decision maker i 's preferences \succ ; under these axioms there exists a closed and convex set Π of probability distributions on S with the property that

$$j \succ k \quad \text{if and only if} \quad E_{\pi}[u(j)] > E_{\pi}[u(k)] \text{ for all } \pi \in \Pi$$

If the inequality changes sign for different probability distributions the two alternatives are not comparable. This is not a special case of the interval order we studied in the previous section. Although all the possible values of $E_{\pi}[u_i(j)]$ describe an interval as before, comparisons are made one probability distribution at the time and not by looking at the extremes of that interval; two alternatives could be ranked even if the corresponding intervals overlap. Despite this difference, the results of the previous section can be applied here by taking advantage of some simple algebra.

The expression above can be rewritten in terms of the expected value of the utility difference between the two alternatives as:

$$j \succ k \quad \text{if and only if} \quad E_{\pi}[u(j) - u(k)] > 0 \text{ for all } \pi \in \Pi$$

and therefore one obtains

$$j \succ k \quad \text{if and only if} \quad E_{\pi^{min}}[u(j) - u(k)] > 0$$

and

$$k \succ j \quad \text{if and only if} \quad E_{\pi^{max}}[u(j) - u(k)] < 0,$$

where we defined $\pi^{min} \in \arg \min_{\pi \in \Pi} E_{\pi}[u(j) - u(k)]$ and $\pi^{max} \in \arg \max_{\pi \in \Pi} E_{\pi}[u(j) - u(k)]$. This last two inequalities can now be used

Suppose the choice set C contains only alternatives j and k . Then, the decision maker's choices can be described by the following set

$$Y = \begin{cases} \{j\} & \text{if } E_{\pi^{min}}[u(j) - u(k)] > 0 \\ \{j, k\} & 0 \in [E_{\pi^{min}}[u(j) - u(k)], E_{\pi^{max}}[u(j) - u(k)]] \\ \{k\} & \text{if } E_{\pi^{max}}[u(j) - u(k)] < 0 \end{cases} .$$

At this point, the same logic of the previous set can be applied to try and identify properties of the probability distribution of $E_{\pi^{min}} [u(j) - u(k)]$ and $E_{\pi^{max}} [u(j) - u(k)]$. Before illustrating this, however, note that individuals' choices will not reveal details about the distribution of Π and $u(\cdot)$ (the preference parameters) beyond what one may learn from the two extreme expectations.

From now on, think of $E_{\pi^{min}} [u(j) - u(k)]$ and $E_{\pi^{max}} [u(j) - u(k)]$ as two random variables which obviously satisfy $\Pr(E_{\pi^{min}} [u(j) - u(k)] \leq E_{\pi^{max}} [u(j) - u(k)]) = 1$. In other words, while consumers may differ in either their utility function for each alternative or in the probability they assign to each state of the world, the only heterogeneity data can be informative about is the one of these extreme expected values. As we observe individuals' choice that follow the rule illustrated by the set Y , we use these observations to learn about some features of the distribution of these random variables. By applying Artstein's inequality here, we observe the following. When $K = \{j\}$, we have

$$C_Y(\{j\}) = \Pr(Y \subset \{j\}) = \Pr(E_{\pi^{min}} [u(j) - u(k)] > 0),$$

while when $K = \{k\}$ we obtain

$$C_Y(\{k\}) = \Pr(Y \subset \{k\}) = \Pr(E_{\pi^{max}} [u(j) - u(k)] < 0).$$

Therefore, from Artstein inequality we conclude that

$$\Pr(y = j) \geq \Pr(E_{\pi^{min}} [u(j) - u(k)] > 0)$$

and

$$\Pr(y = k) \geq \Pr(E_{\pi^{max}} [u(j) - u(k)] < 0)$$

Using the fact that $\Pr(y = j) = 1 - \Pr(y = k)$, we can conclude that

$$\Pr(E_{\pi^{min}} [u(j) - u(k)] > 0) \leq \Pr(y = j) \leq \Pr(E_{\pi^{max}} [u(j) - u(k)] < 0).$$

As before, if the data identifies the values of p_k and p_j as the probability that a randomly chosen individual will choose k and j respectively, the inequalities above imply that

$$\Pr(E_{\pi^{min}} [u(j) - u(k)] > 0) \leq p_j \leq \Pr(E_{\pi^{max}} [u(j) - u(k)] < 0).$$

Thus, knowledge about the probability that a randomly chosen individual will choose j implies restrictions on the marginal distributions of $E_{\pi^{min}} [u(j) - u(k)]$ and $E_{\pi^{max}} [u(j) - u(k)]$. In particular, the lowest expected utility difference between j and k cannot be smaller than the probability of choosing j , and the largest expected utility difference between j and k cannot be larger than the probability of choosing j . So, the inequality above says that data can impose restrictions on preference parameters even when the underlying preferences are not complete.

6.1 Epsilon-contamination model

We now specialize the model of decision making under uncertainty so that we can learn more details about preferences from the partially identified region for the probability distribution of the highest and lowest expectations. In particular, we assume that S has only two elements (denoted H and L) and that $\Pi = \{(1 - \varepsilon)\hat{\pi} + \varepsilon\pi : \pi \in \Delta^2\}$ where Δ^2 is the two dimensional simplex and ε is a random variable that takes values in the interval $[0, 1]$. This is known as the ε -contamination model as the set Π is constructed by “mixing” a reference distribution with arbitrary distributions in the simplex. One can think of it as building a probabilistic neighborhood of size epsilon around the reference measure. The size of ε is then a measure of the amount of ambiguity.

Consumers face two choices $\{j, k\}$, and for simplicity we assume that $u(j) = 0$ in all states of the world.¹⁵ For option k , however, we assume the utility of each consumer can be either u_L or u_H , with $u_L < 0 < u_H$, depending on the state of the world. The reference probability distribution is $\hat{\pi} = (\hat{\pi}_L, \hat{\pi}_H)$. By construction, the probabilities assigned to state L satisfy

$$(1 - \varepsilon) \hat{\pi}_L \leq \pi_L \leq (1 - \varepsilon) \hat{\pi}_L + \varepsilon$$

The bounds on the probability assigned to state L induce bounds on the expected utility from choosing option k ; in particular,

$$\begin{aligned} E_{\hat{\pi}}[u(k)] &= \hat{\pi}_L u_L + \hat{\pi}_H u_H + (\pi_L - \hat{\pi}_L) \varepsilon u_L + (\pi_H - \hat{\pi}_H) \varepsilon u_H \\ &= \bar{U} + \varepsilon [(\pi_L - \hat{\pi}_L) u_L + (\pi_H - \hat{\pi}_H) u_H] \\ &= \bar{U} + \varepsilon [\pi_L u_L + \pi_H u_H - \bar{U}] \end{aligned}$$

where we use \bar{U} to denote the expected utility of alternative k according to the reference measure ($\bar{U} = \hat{\pi}_L u_L + \hat{\pi}_H u_H$). Therefore, since $u_L < 0 < u_H$, the highest and lowest expected utility are given by

$$\bar{U} + \varepsilon (u_H - \bar{U}) \quad \text{and} \quad \bar{U} + \varepsilon (u_L - \bar{U})$$

respectively.

We can now use the inequalities implied by Artstein’s inequality (here we use $\bar{U} - \varepsilon (\bar{U} - u_L)$ instead of V_L , and $\bar{U} + \varepsilon (u_H - \bar{U})$ instead of V_U). to derive that

$$\Pr(\bar{U} - \varepsilon (\bar{U} - u_L) < 0) \geq p_0 \geq \Pr(\bar{U} + \varepsilon (u_H - \bar{U}) < 0).$$

These inequalities imply inequalities on the CDF of ε . With some algebra we can show that

$$\Pr(\bar{U} - \varepsilon (\bar{U} - u_L) < 0) = \Pr\left(\varepsilon > \frac{\bar{U}}{(\bar{U} - u_L)}\right) \quad (19)$$

¹⁵By setting the utility of one option constant, we collapse reduce the model of this section to the one of the previous section, in that the upper and lower bounds on the expected utility of the uncertain options determine behavior.

and

$$\Pr(\bar{U} + \varepsilon(u_H - \bar{U}) < 0) = \Pr\left(\varepsilon < -\frac{\bar{U}}{(u_H - \bar{U})}\right). \quad (20)$$

By definition, we have that $u_L \leq \bar{U} \leq u_H$. If $\bar{U} > 0$, then the constraint (20) is non binding since $\Pr(\bar{U} + \varepsilon(u_H - \bar{U}) < 0)$ is trivially zero. On the other hand if $\bar{U} < 0$, then the constraint in (19) is non binding since $\Pr(\bar{U} - \varepsilon(\bar{U} - u_L) < 0)$ is trivially zero.

Let $\theta = (\hat{\pi}, u_L, u_H, F_\varepsilon) \in \Theta = \Delta^2 \times \mathbb{R}_- \times \mathbb{R}_+ \times \Gamma$ be the parameter space where Γ is the set of all random variables on $[0, 1]$. Using the definition of the identification region in (16) we can write the identification region of the parameter θ . It is easy to see that Θ is too high dimensional and the identification region Θ_I in (16) is hard to characterize. At this point it is useful to fix some of the dimensions and see what information can be drawn on the remaining dimensions. We explore some of these options below by assuming that the utility of option k in each state as well as the probability distribution $\hat{\pi}$ are the same for all consumer. In this case, the only source of heterogeneity is given by differences in the perception of ambiguity.

Example. Suppose $\hat{\pi}, u_L$ and u_H are fixed and ε is distributed $B(\alpha, \beta)$ for unknown α and β . The identification region is all the pairs of α and β such that (19) and (20) are satisfied. Figure (??) shows the possible pairs of (α, β) for $\hat{\pi} = (0.6, 0.4)$, $u_L = -2$, $u_H = 1$ and $p_0 = 0.37$.

Moreover, we can compute the identification rage for the average and variance of ε . In the example above, the Beta distributions admitted into the identification region have an expectation ranging from 0.55 to 1 and a variance of at most 0.22. These numbers tell us about the degree of ambiguity faced by the agents as suggested by the data. Intuitively, the degree of ambiguity is high enough to offset the negative expected value of alternative k .

7 Voting and Abstaining

Until this point we assumed that agents have to make a choice from the set of alternatives \mathcal{A} . The important part of this assumption is that a utility is assigned to each of the elements in the set \mathcal{A} , potentially an interval. It is possible that \mathcal{A} contains a 'no action' alternative (e.g. do not buy any car) but this alternative gives a certain utility to the DM which is often normalized to zero. There are situations where the DM can abstain from making any choice. The most prominent example is voting. When a DM (voter) comes to the polling place she can chose to abstain from voting on some or all of the races on the ballot. The utility associated with this choice is undefined. Denote the extended alternatives set as,

$$\bar{\mathcal{A}} = \mathcal{A} \cup \{\emptyset\}.$$

In the case of voting, by abstaining the DM delegates the decision to other DMs.

7.1 Voting in Ohio

We apply the theory described above to voting in Ohio. The voting in Ohio presents a neat IV example. It is known that the order of the candidates on the ballot affects the chances of candidates to be elected. As a result ordering the candidates in alphabetical order may bias the results in favor of candidates with certain family names.

To be completed

8 Conclusions

TBA

Appendices

A Technical Results For Section 3.1

To demonstrate the effect of Artstein's Lemma we prove the following two claims related to the sharpness of the identification set \mathcal{P}^I defined in equation (3) in Section 3.1. First we show that \mathcal{P}^I is a strict subset of \mathcal{P} under mild conditions. We then show that all the inequality implied by Artstein's Lemma are potentially binding.

Claim A.1 \mathcal{P}^I is a strict subset of \mathcal{P} if $\exists a \in A$ such that $1 > \Pr(y = a) > 0$.

Proof. There is $a' \neq a$ such that $\Pr(y = a') > 0$. Therefore, $\Pr(\underline{u}(a) > \bar{u}(a')) < 1$. This is a restriction on the joint distribution of $\{\underline{u}(a), \bar{u}(a)\}_{a \in A}$ that is not included in the definition of \mathcal{P} . Therefore, \mathcal{P}^I is a strict subset of \mathcal{P} . ■

We now want to demonstrate that the conditions in Artstein's Lemma essential in the sense that using less restrictions yields a strictly larger set than the identification region. Let

$$\mathcal{P}^1 = \{P \in \mathcal{P} : C_M(\{a\}) \leq \Pr(y = a) \forall a \in A\}$$

be the set of joint distributions that satisfy Artstein's Lemma only for subsets of $A \subset \mathcal{A}$ of cardinality 1.

Claim A.2 If $|\mathcal{A}| > 2$, $\mathcal{P}^I \subsetneq \mathcal{P}^1$.

Proof. Since $|\mathcal{A}| > 2$ and $|A| = 1$, there are two alternatives $a, a' \in \mathcal{A}$ such that $a \neq a'$ and $\{a, a'\} \neq A$. By Artstein's Lemma $C_M(\{a\}) \leq \Pr(y = a)$, $C_M(a') \leq \Pr(y = a')$ and $C_M(\{a, a'\}) \leq \Pr(y \in \{a, a'\})$. The first two inequalities yield $C_M(\{a\}) + C_M(\{a'\}) \leq \Pr(y = a) + \Pr(y = a') = \Pr(y \in \{a, a'\})$ where the last equality is implied by $a \neq a'$. This inequality is included in the definition of \mathcal{P}^1 . Since $C_M(\{a\}) + C_M(\{a'\}) \leq C_M(\{a, a'\})$, the inequality $C_M(\{a, a'\}) \leq \Pr(y \in \{a, a'\})$ may impose an additional restriction on the distribution of y and therefore on the joint distribution of $\{\underline{u}(a), \bar{u}(a)\}_{a \in A}$. ■

B Incomplete Probit Example

Consider the following binary choice model. Let $\mathcal{A} = \{0, 1\}$ be the alternatives set. Normalize the utility from choice $u_i(0) = 0$ for all i . The utility from choice 1 is vague. we assume that $u_i(1) \in [\beta_i, \beta_i + \delta_i]$ where $\delta_i > 0$ and the joint distribution of (β, δ) is unknown. From Artstein's inequalities we know that

$$\Pr(\beta + \delta < 0) \leq p_0 \leq \Pr(\beta < 0).$$

These two inequalities impose a certain restrictions on the joint distribution of (β, δ) but these restrictions are limited.

B.1 One dimensional heterogeneity

Suppose now that $\beta_i = \beta - \varepsilon_i$ where $\varepsilon_i \sim F_\varepsilon$ for a known distribution F_ε (e.g. a logistic or standard normal distribution). δ is assumed to be constant across all decision maker. In other words, there is only one source of heterogeneity. This implies that the following is an equivalent model. Set $u_i(0) = \varepsilon_i$ with $\varepsilon_i \sim F_\varepsilon$ and $u_i(1) = \beta + \delta$ constant for all i .

Under these assumptions, $\Pr(\beta < 0) = 1 - F_\varepsilon(\beta)$ and $\Pr(\beta + \delta < 0) = 1 - F_\varepsilon(\beta + \delta)$. Since $p_0 = 1 - p_1$, we can write

$$F_\varepsilon(\beta + \delta) \leq p_1 \leq F_\varepsilon(\beta),$$

or

$$\beta \leq F_\varepsilon^{-1}(p_1) \leq \beta + \delta.$$

The identification region is, therefore,

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

As before, suppose there is an additional variable z independent of (β, δ) with support \mathcal{Z} . In that case,

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

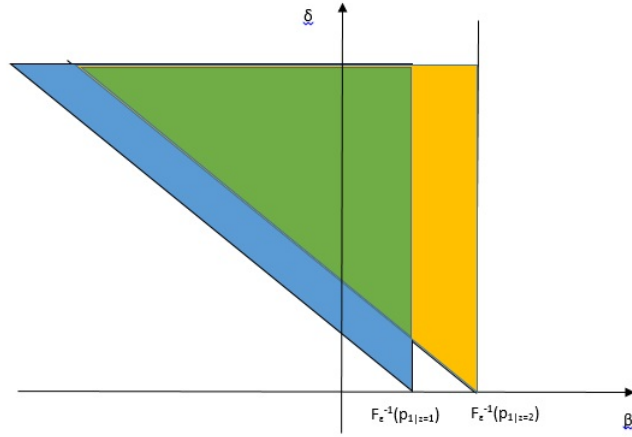


Figure 4: One Dimensional Heterogeneity

B.2 Two dimensional heterogeneity

Here we add an additional source of heterogeneity - the amount of vagueness. Let $\sigma_i = \sigma - \eta_i$ for each i . Using the same logic as before, we get

$$\Pr(\varepsilon_i \leq \beta) \leq p_1 \leq \Pr(\varepsilon_i + \eta_i \leq \beta + \delta).$$

These inequalities translate to

$$\begin{aligned}\beta &\leq F_{\varepsilon}^{-1}(p_1) \\ F_{\varepsilon+\eta}^{-1}(p_1) &\leq \beta + \delta.\end{aligned}$$

The identification set is defined accordingly. Suppose $\beta_i \sim N(\beta, \sigma_{\beta}^2)$ and $\delta_i \sim \text{Exp}(\frac{1}{\delta})$ independent of β_i . Collecting the parameters we have $\theta = (\beta, \sigma_{\beta}^2, \delta)$. Note that $\beta_i + \delta_i$ distribution is called Exponentially Modified Gaussian (EMG) distribution.¹⁶

¹⁶Let $\lambda = \frac{1}{\delta}$, then the CDF of an EMG distribution is given by $EMG(z; \beta, \sigma_{\beta}^2, \lambda) = \frac{1}{2} [1 - q_1 \cdot q_2 + q_3]$ where $q_1 = \exp\left[\lambda\left(\frac{\lambda\sigma_{\beta}^2}{2} + \mu + z\right)\right]$, $q_2 = 1 - \text{erfc}\left(\frac{\sigma_{\beta}}{\sqrt{2}}\left(\lambda + \frac{\beta - z}{\sigma_{\beta}^2}\right)\right)$, $q_3 = \text{erfc}\left(\frac{1}{\sqrt{2}\sigma_{\beta}}(z - \mu)\right)$ and $\text{erfc}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x^2) dx$ is the error function.

References

- Robert J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30:445–462, 1962.
- Truman F. Bewley. Knightian decision theory: Part i. Technical report, Cowles Foundation, 1986.
- Truman F. Bewley. Knightian decision theory: Part i. *Decisions in Economics and Finance*, 2:79–110, 2002.
- Juan Dubra, Fabio Maccheroni, and Efe A. Ok. Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115:118–133, 2004.
- Efe A. Ok. Utility representation of an incomplete preference relation. *Journal of Economic Theory*, 104:429–449, 2002.
- Luca Rigotti and Chris Shannon. Uncertainty and risk in financial markets. *Econometrica*, 73:203–243, 2005.
- Giuseppe Lopomo, Luca Rigotti, and Chris Shannon. Knightian uncertainty and moral hazard. *Journal of Economic Theory*, 146:1148–1172, 2012.
- Peter C. Fishburn. *Utility Theory For Decision Making*. Wiley and Sons, 1970.
- Ilya Molchanov. *Theory of Random Sets*. Springer, 2005.
- Arie Beresteanu, Ilya Molchanov, and Francesca Molinari. Partial identification using random set theory. *Journal of Econometrics*, 166:17–32, 2012.
- Arie Beresteanu, Ilya Molchanov, and Francesca Molinari. Sharp identification regions in models with convex moment predictions. *Econometrica*, 79:1785–1821, 2011.
- Masuo Hukuhara. Integration des applications mesurables dont la valeur est un compact convexe. *Funkcialaj Ekvacioj*, 10:205–223, 1967.
- Aviv Nevo and Adam Rosen. Identification with imperfect instruments. *Review of Economics and Statistics*, 93(3):695–671, 2012.