RANDOM UTILITY MODELS
WITH ORDERED TYPES AND DOMAINS

JOSE APESTEGUIA† AND MIGUEL A. BALLESTER‡

ABSTRACT. Random utility models in which heterogeneity of preferences is modeled by means of an ordered collection of utilities, or types, provide a powerful framework for the understanding of a variety of economic behaviors. However, previous theoretical treatments are very demanding in terms of the data they require. This paper studies the theoretical foundations of ordered random utility models with the fundamental objective of meeting standard empirical requirements. This is done by working with arbitrary domains containing decision problems ordered by the structure of types. The model is characterized by means of two simple properties. The generality and applicability of the setting is demonstrated by means of a series of extensions and a particularization for decisions under risk. A goodness-of-fit measure is proposed for the model and proof is also provided of the strong consistency of extremum estimators defined upon it. The paper concludes with the application of the model to a dataset on lottery choices.

Keywords: Random utility model; Ordered type-dependent utilities; Arbitrary domains; Non-parametric; Goodness-of-fit; Extremum estimators; Decision under risk.
JEL classification numbers: C00; D00

Date: July, 2020.

*We thank Jay Lu and Christoph Semken for helpful comments and Angelo Gutierrez for outstanding research assistance. Financial support by the Spanish Ministry of Science (PGC2018-098949-B-I00) and Balliol College is gratefully acknowledged.

†ICREA, Universitat Pompeu Fabra and Barcelona GSE. E-mail: `jose.apesteguia@upf.edu`.

‡University of Oxford. E-mail: `miguel.ballester@economics.ox.ac.uk`.

1. Introduction

Ordered collections of utilities, or types, provide a handy tool for the analysis of behavioral traits, such as risk aversion, delay aversion or altruism, which can be intuitively ordered. Interestingly, the ordered structure of types enables some alternatives to be considered *higher than* others, simply because they are preferred by higher types, as occurs under the notions of safer lottery, earlier stream of payoffs, or more altruistic distribution. Decision problems composed of alternatives ordered by the "higher than" principle are very informative because, in such problems, the lowest types choose one alternative, and successively higher types choose successively higher alternatives. Intuitively, this structure facilitates the revelation of the behavioral trait at stake, thus often constituting one of the desiderata for a good empirical setting or experimental design.

Sound theoretical foundations for random utility models that introduce heterogeneity using an ordered collection of types are therefore imperative. Unfortunately, previous theoretical work impose demanding domain assumptions, such as that the grand set of alternatives is completely ordered by the "higher than" principle, or that choice data on every possible subset of alternatives are available. These requirements are rarely met in practice. The fundamental objective of this paper is to bridge the gap between the theory of random utility models with ordered types and the standard empirical demands for their actual implementation. We do so by (i) requiring that decision problems for which data exist are ordered, instead of imposing a complete ordering of the grand set of alternatives, and (ii) assuming that data are available on an arbitrary domain of such decision problems, thus eliminating the need for particularly demanding domains. To further enhance applicability, heterogeneity is fully non-parametrically treated throughout; that is, without imposing any particular probability distribution over the type space.

The first result of the paper provides a characterization of the choice frequencies that can be generated by any random utility model over a given space of ordered types $\mathcal{T}$. Two properties are used. The first, $\mathcal{T}$-Extremeness, is a well-known property of random utility models which states that only those alternatives that are maximal for at least one type in the population can be chosen with strictly positive probability. The second property, $\mathcal{T}$-Monotonicity, also emanates directly from the structure of random utility models. In a nutshell, suppose that the set of types leading to alternatives

B_1, within decision problem A_1, is a subset of the set of types yielding choices in B_2, within decision problem A_2. Then, $\mathcal{T}$-Monotonicity states that the cumulated choice frequency of alternatives in B_1, within decision problem A_1, must be smaller than that of alternatives in B_2, within decision problem A_2. Interestingly, this paper shows that, when decision problems are ordered, these two simple properties are not only necessary but also sufficient. Importantly, it is noted that the proof of the characterization result is constructive and that the model is uniquely identified at all thresholds at which the maximal alternative of some decision problem changes.

Further elaboration on two generalizations of the domain assumption lending even greater applicability to our framework is then provided. First, it is noticed that our analysis goes through even if the alternatives of a decision problem are not necessarily ordered by the "higher than" notion, as long as every alternative in every menu is chosen by an interval of types. Secondly, it is shown that any decision problem can belong to the domain, as long as it satisfies a simple richness condition; namely, that one ordered decision problem in the domain yields a finer collection of type intervals than that yielded by the unstructured decision problem. These two results significantly expand the potential use of random utility models over ordered collections of types.

The basic results are then illustrated by particularizing the analysis to decisions under risk. In this setting, the most prominent type-dependent utilities are expected utilities, ordered by increasing levels of risk aversion, thus generating a riskiness/safety order over lotteries. A variety of decision problems, representative of the experimental designs in the literature, are then shown to be, in fact, ordered and therefore suitable for theoretical analysis using the tools proposed in this paper. These are decision problems in which lotteries: (i) are formed by two states with fixed probabilities, as in multiple price lists or convex budget set experiments, (ii) are defined on the same Marschak-Machina triangle, i.e., involving the same three monetary payoffs or (iii) award a (possibly different) prize with a (possibly different) probability. Moreover, and in relation to the second domain generalization mentioned above, it is shown that the richness condition can be met by using elementary decision problems, as it is always possible to use lotteries from any given Marschak-Machina triangle.

The first part of the paper concludes with an analysis of theoretical variations of the model that are useful for addressing other empirical questions. We begin by introducing a tremble version of the model, in which choices of alternatives that are never maximal are observed. We then discuss the case in which the analyst has choice information that potentially varies across subpopulations, as occurs with control-treatment studies or gender/age-specific data. Finally, we consider a setting involving infinite type spaces and decision problems. In all three of these cases, our simple properties, $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity, are the basis for understanding the theoretical structure of these variations.

Given our interest in connecting theory with empirics, we then present results that are of econometric interest. Considering the case of finite data, which, due to sampling issues, may violate $\mathcal{T}$-Monotonicity, we lay down an intuitive goodness-of-fit measure based on the possibility that the underlying distribution of types is perturbed at each decision problem. This measure is the minimum perturbation required to explain all observed choice frequencies, and implicitly defines an extremum estimator. Most importantly, subsequent analysis shows that any estimator in this class is strongly consistent. That is, as the number of observations per choice problem increases, the estimator converges to the true probability distribution over types.

We conclude with an empirical illustration of our theoretical results, using an existing experimental dataset involving decision problems over lotteries. We use the ordered collection of types formed by CRRA expected utilities and argue that the experimental domain is ordered. We then estimate the ordered random utility model using maximum likelihood, and briefly discuss some findings.

2. Related Literature

There is growing theoretical, econometric and empirical interest in random utility models using ordered collections of types. The closest paper to this one is Apesteguia, Ballester and Lu (2017), which provides a theoretical analysis of the single-crossing random utility model (SCRUM).[1] The fundamental difference between the two papers is that the earlier one deals with the characterization of the model using a traditional stochastic choice theory approach, while the present one is of a more practical nature, and aimed at responding to the theoretical and econometric challenges posed by the

[1] The single-crossing property has been intensively studied in economics at least since Mirrlees (1971) and Spence (1974), and has had a large impact in the social, biological and health sciences (see, e.g., Greene and Hensher, 2010).

empirical implementation of the model.[2] This leads to four main differences. Firstly, this paper models the conventional approach used in applied work, where the analyst confronts the situation with a favored, fixed, family of ordered types. This lends content to the properties used in our characterization.[3] Secondly, whereas SCRUMs impose the single-crossing condition on the grand set of alternatives, X, the only condition here is that completeness holds within each menu, and not necessarily on X.[4] Thirdly, whereas the characterization of SCRUMs uses a universal domain; i.e., it requires data on every single subset of the grand set of alternatives, here we work with arbitrary domains. It is thanks to these two last differences that the present paper satisfies the empirical requirements for implementation. Both the completeness on X and the universal choice domain are assumptions that facilitate the theoretical treatment of the model, but are very rarely met in practice. With the relaxation of these two assumptions, the model meets the majority of existing datasets. Finally, in line with the practical motivation for this study, the findings enable us to derive econometric results of direct interest for the implementation of the model. In particular, we introduce a class of estimators which are shown to be strongly consistent. All in all, the present paper takes important steps towards endowing ordered random utility models with full-fledged empirical content.

In a recent theoretical contribution, Filiz-Ozbay and Masatlioglu (2020) study a random model using an ordered collection of choice functions rather than utilities and thus, importantly, provide the theoretical foundations for what can be considered a model of stochastic, boundedly rational choice. The main difference between their paper and ours is that we work on the practical implementation of random utility models.

A handful of recent empirical papers focus on trying to exploit the single-crossing condition. Barseghyan, Molinari and Thirkettle (2019) use random utility models satisfying the single-crossing condition to provide a semi-parametric identification of attention models under risk taking. Chiappori, Salanié, Salanié and Gandhi (2019)

[2]The stochastic choice theory literature is turning to the issue of the empirical implementation of the theoretical models developed in the field. For example, Dardanoni, Manzini, Mariotti and Tyson (2020) study limited-attention stochastic choice models, where the choice domain involves a single menu of alternatives, and Cattaneo, Ma, Masatlioglu and Suleymanov (2020) establish non-parametric identification results for their random attention model.

[3]In a similar fashion, Gul and Pesendorfer's study (2006) of random expected utility uses an extremeness property defined on the basis of the class of expected utilities.

[4]Indeed, in Section 4 we significantly relax this already mild assumption.

also impose the single-crossing condition on individual risk preferences in a parimutuel horse-racing setting to establish the equilibrium conditions and ultimately identify the model. Our paper provides theoretical foundations for the model and its estimation in general settings, beyond that of decisions under risk, while using arbitrary domains.

A series of applied papers have implemented parametric versions of the random utility model, over an ordered collection of types, to estimate a specific behavioral trait; most frequently, risk aversion.[5] Barsky, Juster, Kimball and Shapiro (1997) is one of the first examples of the use of this methodology, where the ordered structure of a decision problem involving lotteries is exploited to obtain population estimates of risk aversion and perform covariate analysis. Cohen and Einav (2007) use data on auto insurance contracts, showing that any given probability of accident leads to an ordered decision problem of deductibles and premiums, thereby facilitating the estimation of risk aversion. Andersson, Holm, Tyran and Wengström (2018) use decision problems involving two states with fixed probabilities to show that choice variability is determined by cognitive ability rather than risk aversion. Our paper contributes to this applied literature by providing foundations for a general, non-parametric, version of the model.

The econometrics literature on the non-parametric identification of ordered discrete choice models is also of relevance here (see Cunha, Heckman and Navarro (2007), and references cited therein, and Greene and Hensher (2010) for a survey). The papers in this literature typically involve a single menu of ordered alternatives and the identification argument relies on relating the probability of choice of any alternative with the mass of types for which the alternative is optimal, which, given the structure, forms an interval. Intuitively, we use similar techniques for our results. Importantly, notice that the use of a unique menu precludes the possibility of testing the model. However, by extending the model to arbitrary domains of ordered menus, we are able to provide characterizing conditions for ordered random utility models.

[5]See Coller and Williams (1999) and Warner and Pleeter (2001) for similar estimation exercises within the context of time preferences, or Apesteguia, Ballester and Gutierrez (2020) and Jagelka (2020) for joint estimations of risk and time preferences.

3. Ordered Random Utility Models and Ordered Domains

Let X be the set of all alternatives. We fix an ordered collection of type-dependent utilities $\{U_t\}_{t\in\mathcal{T}}$, where $\mathcal{T} = \{1, 2, \ldots, T\}$.[6] Since all of our results are ordinal in nature, no parametric assumption is required here; we could equivalently work with the corresponding collection of ordinal preferences. A random utility model over $\mathcal{T}$, or $\mathcal{T}$-RUM, is defined by a probability distribution ψ over $\mathcal{T}$, which describes the probability mass with which each type is realized.[7] In a decision problem, one of the utility functions is realized and maximized, thus determining the choice. Decision problems, or menus, are finite subsets of alternatives. We work with an arbitrary collection of menus, $\{A_j\}_{j\in\mathcal{J}}$, where $\mathcal{J} = \{1, 2, \ldots, J\}$. Given the $\mathcal{T}$-RUM with distribution ψ, the probability of choice of alternative x in menu j is $\psi(\mathcal{T}(x, j))$, where $\mathcal{T}(x, j)$ denotes the set of types for which alternative x is the utility maximizer in menu j.

Ordered collections of utilities induce an order over some pairs of alternatives. We say that alternative x_h is *higher than* alternative x_l whenever there exists $t^* \in \mathcal{T} \setminus \{T\}$ such that $U_t(x_l) > U_t(x_h) \Leftrightarrow t \leq t^*$. In this case we write $x_l \lhd x_h$, and, as usual, $x \unlhd y$ whenever $x \lhd y$ or $x = y$. In words, x_h is higher than x_l if x_h is the preferred alternative of high types (with at least type T expressing this preference) and x_l is the preferred alternative of low types (with at least type 1 expressing this opposite preference). For instance, types can be ordered by risk aversion, delay aversion or altruism, and hence the notion of a higher alternative corresponds to the notions of a safer lottery, a less delayed stream of payoffs, or a more altruistic distribution. We now introduce the only relevant restriction in the paper: i.e., that every menu in the domain is ordered, in the sense that its maximal alternatives can be ordered by $\lhd$.

Domain of Ordered Menus. For every $j \in \mathcal{J}$, $\lhd$ is complete over $\{x : \mathcal{T}(x, j) \neq \emptyset\}$.

Domains composed by ordered menus are pervasive when studying a particular behavioral trait.

---

[6] Since we present our basic results on a finite domain, finiteness of types can be assumed without loss of generality. The extension to infinite type spaces is discussed in Section 6. For ease of exposition, we also assume no role for indifferences, i.e., utility functions are strict over the domain for which data exist.

[7] We denote by Ψ the set of all such distributions.

3.1. **A characterization of $\mathcal{T}$-RUMs.** Suppose that the analyst has access to ideal (infinite) data in the form of a stochastic choice function p, which is simply a map from $X \times \mathcal{J}$ to $[0,1]$ such that, for every $j \in \mathcal{J}$, $p(x,j) > 0$ implies that $x \in A_j$, and $\sum_{x \in A_j} p(x,j) = 1$. A first property of $\mathcal{T}$-RUMs, which we name as in Gul and Pesendorfer (2006), reads as follows:

$\mathcal{T}$-Extremeness: $p(x,j) > 0 \Rightarrow \mathcal{T}(x,j) \neq \emptyset$.

Clearly, $\mathcal{T}$-Extremeness is a result of the optimizing nature of RUMs. It implies that only those alternatives that are maximal for some type can be observed with strictly positive mass. The second property reads as:

$\mathcal{T}$-Monotonicity: $\bigcup_{x \in B} \mathcal{T}(x,j) \subseteq \bigcup_{x' \in B'} \mathcal{T}(x',j') \Rightarrow \sum_{x \in B} p(x,j) \leq \sum_{x' \in B'} p(x',j')$.

$\mathcal{T}$-Monotonicity is a result of the menu-independent structure of RUMs. Whenever the set of types leading to alternatives B in menu j is contained in the set of types leading to alternatives B' in menu j', the cumulated probability of the former block of alternatives must be lower than that of the latter.

We now show that, when the domain is composed exclusively of ordered menus, these two properties are not only necessary but also sufficient for $\mathcal{T}$-RUMs.

Theorem 1. *In a domain of ordered menus, p satisfies $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity if, and only if, p is a $\mathcal{T}$-RUM.*

Proof of Theorem 1: The necessity of the axioms is evident and thus omitted. We now prove their sufficiency, by proceeding through a series of claims. We then assume that the domain is composed of ordered menus, and that p satisfies $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity.

Claim 1. For every menu $j \in \mathcal{J}$ and alternative $x \in A_j$, $\mathcal{T}(x,j)$ is an interval.

Proof of Claim 1: Suppose, by way of contradiction, that the claim is not true. Let (x,j) be a pair such that types $t_1 < t_2 < t_3$ exist, with $\{t_1, t_3\} \subseteq \mathcal{T}(x,j)$, but $t_2 \notin \mathcal{T}(x,j)$. Let $z \in A_j$ be the alternative for which $t_2 \in \mathcal{T}(z,j)$. From the joint consideration of types t_1 and t_2, it must be that $z \ntriangleleft x$. From the joint consideration of types t_2 and t_3, it follows that $x \ntriangleleft z$. This contradicts the fact that menu j is ordered, and concludes the proof of the claim. $\square$

Claim 2. For every menu $j \in \mathcal{J}$ and alternative $x \in A_j$ such that $\mathcal{T}(x,j) \neq \emptyset$, $\bigcup_{y \in A_j, y \trianglelefteq x} \mathcal{T}(y,j) = \{1, 2, \ldots, \max \mathcal{T}(x,j)\}$.

Proof of Claim 2: Suppose, by way of contradiction, that the claim is not true. Then, there exists t^* such that either: (i) $t^* \leq \max \mathcal{T}(x,j)$ and $t^* \notin \bigcup_{y \in A_j, y \trianglelefteq x} \mathcal{T}(y,j)$, or (ii) $t^* > \max \mathcal{T}(x,j)$ with $t^* \in \bigcup_{y \in A_j, y \trianglelefteq x} \mathcal{T}(y,j)$ hold. In both cases, let $z \in A_j$ be the alternative for which $t^* \in \mathcal{T}(z,j)$. In case (i), we have $z \ntrianglelefteq x$ by assumption, implying that t^* and $\max \mathcal{T}(x,j)$ must be different types. Their joint consideration guarantees, furthermore, that $x \ntriangleleft z$, thus contradicting the fact that menu j is ordered. In case (ii), we know, by assumption, that $z \trianglelefteq x$. Since $t^* > \max \mathcal{T}(x,j)$, it must be that $x \neq z$ and $u_{t^*}(x) > u_{t^*}(z)$, which contradicts the assumption that $t^* \in \mathcal{T}(z,j)$, thus concluding the proof of the claim. □

We now consider the sub-collection of types $\mathcal{T}^I \subseteq \mathcal{T}$ and the mapping $F : \mathcal{T}^I \to [0,1]$ defined by:

$$\mathcal{T}^I = \{t \in \mathcal{T} : \text{ there exists } (x,j) \text{ such that } \max \mathcal{T}(x,j) = t\}.$$

$$F(t) = \sum_{y \in A_j, y \trianglelefteq x} p(y,j) \text{ with } t = \max \mathcal{T}(x,j) \text{ for some } (x,j).$$

Claim 3. F is a single-valued increasing map.

Proof of Claim 3: To see this, consider two types $t_1, t_2 \in \mathcal{T}^I$ such that $t_1 \leq t_2$. By definition of $\mathcal{T}^I$, there exist pairs (x_1, j_1) and (x_2, j_2) such that $t_1 = \max \mathcal{T}(x_1, j_1)$ and $t_2 = \max \mathcal{T}(x_2, j_2)$. By Claim 2, and the fact that $t_1 \leq t_2$, we know that $\bigcup_{y \in A_{j_1} : y \trianglelefteq x_1} \mathcal{T}(y, j_1) = \{1, 2, \ldots, t_1\} \subseteq \{1, 2, \ldots, t_2\} = \bigcup_{y \in A_{j_2} : y \trianglelefteq x_2} \mathcal{T}(y, j_2)$. The use of $\mathcal{T}$-Monotonicity guarantees that $\sum_{y \in A_{j_1}, y \trianglelefteq x_1} p(y, j_1) \leq \sum_{y \in A_{j_2}, y \trianglelefteq x_2} p(y, j_2)$, with equality when $t_1 = t_2$, which proves the claim. □

Claim 4. $T \in \mathcal{T}^I$, with $F(T) = 1$.

Proof of Claim 4: To see the first part, consider any menu j and let x be the alternative such that $T \in \mathcal{T}(x,j)$. It can only be the case that $\max \mathcal{T}(x,j) = T$, and hence, $T \in \mathcal{T}^I$. To see the second part, consider any menu j, and any alternative $y \in A_j$ such that $\mathcal{T}(y,j) \neq \emptyset$. Since menu j is ordered, the joint consideration of types $\max \mathcal{T}(y,j)$ and $\max \mathcal{T}(x,j) = T$ guarantees that $y \trianglelefteq x$. Hence, the use of $\mathcal{T}$-Extremeness and the definition of stochastic choice function guarantee that $1 =$

$\sum_{y\in A_j} p(y,j) \geq \sum_{y\in A_j, y\trianglelefteq x} p(y,j) = F(t) \geq \sum_{y\in A_j:\mathcal{T}(y,j)\neq\emptyset} p(y,j) = 1$, which proves the claim. □

Using Claims 3 and 4, we are able to construct a map $G : \mathcal{T} \cup \{0\} \to [0,1]$ that extends F, is weakly increasing, and gives $G(0) = 0$ and $G(1) = 1$. Trivially, $\psi(t) = G(t) - G(t-1)$ is a probability distribution over $\mathcal{T}$. We then consider the $\mathcal{T}$-RUM defined by ψ.

Claim 5. For every menu $j \in \mathcal{J}$, and alternative $x \in A_j$, $p(x,j) = \psi(\mathcal{T}(x,j))$.

Proof of Claim 5: If $\mathcal{T}(x,j) = \emptyset$, we know that $\mathcal{T}$-Extremeness guarantees that $p(x,j) = 0$, which is precisely the probability assigned by the $\mathcal{T}$-RUM. Whenever $\mathcal{T}(x,j) \neq \emptyset$, Claim 1 guarantees that $\mathcal{T}(x,j)$ is an interval. If $1 \in \mathcal{T}(x,j)$, we know, by construction, that $p(x,j) = F(\max\mathcal{T}(x,j)) = G(\max\mathcal{T}(x,j)) = \psi(\mathcal{T}(x,j))$, as desired. If $1 \notin \mathcal{T}(x,j)$, let z be the highest alternative in A_j, according to $\triangleleft$, satisfying $\mathcal{T}(z,j) \neq \emptyset$ and $z \triangleleft x$. It must obviously be the case that $\mathcal{T}(x,j) = \{\max\mathcal{T}(z,j) + 1, \max\mathcal{T}(z,j) + 2, \ldots, \max\mathcal{T}(x,j)\}$ and, by construction, $p(x,j) = F(\max\mathcal{T}(x,j)) - F(\max\mathcal{T}(z,j)) = G(\max\mathcal{T}(x,j)) - G(\max\mathcal{T}(z,j)) = \psi(\mathcal{T}(x,j))$. This proves the claim. □

Having constructed a $\mathcal{T}$-RUM that rationalizes all choice probabilities, we have proved the theorem. ■

The intuition of the proof is as follows. The ordered structure of menus guarantees that the set of types $\mathcal{T}(x,j)$ is always an interval and, consequently, $\bigcup_{y\in A_j, y\trianglelefteq x} \mathcal{T}(y,j)$ is always of the form $\{1, 2, \ldots, \max\mathcal{T}(x,j)\}$. Hence, the map given by $F(\max\mathcal{T}(x,j)) = \sum_{y\in A_j, y\trianglelefteq x} p(y,j)$ is single-valued, is weakly increasing and satisfies $F(T) = 1$. We can then construct a monotone extension G of this map over the entire collection of types, with $G(T) = 1$, thereby forming the CDF of a probability distribution ψ that can be shown to rationalize the data.

Importantly, the above construction starts with types t for which there exists a menu j and $x \in A_j$ with $\max\mathcal{T}(x,j) = t$. These types, which we denote by $\mathcal{T}^I$, are the set of identifiable types, i.e., the CDF of the distribution of types can always be identified for these types.[8] The maximum number of types that can always be identified is

[8]Notice that whenever $t \notin \mathcal{T}^I$, the CDF at t cannot generally be identified. To see this, consider e.g. $t_1 < t < t_2$, where t_1 and t_2 are two consecutive types in $\mathcal{T}^I$. If the value of the CDF at t_2

$\sum_{j=1}^{J} |\{x \in A_j : \mathcal{T}(x,j) \neq \emptyset\}| - J$, a bound that is reached if every pair (x,j) such that $\mathcal{T}(x,j) \neq \emptyset$ generates a different type $\max \mathcal{T}(x,j)$. We can say that the model is fully identified whenever, for every $t \in \mathcal{T}$, there exists a menu j and $x \in A_j$ such that $\max \mathcal{T}(x,j) = t$, which occurs, for instance, whenever, for every $t \in \mathcal{T}$, the domain contains a binary menu $\{x,y\}$, with $x \triangleleft y$, such that type t is the last type to choose x over y. In an experimental setting, the analyst may obviously use these results in order to select the domain that allows identification of the most important components of the CDF, while keeping the experiment as simple as possible.

4. Beyond ordered menus

Domains of ordered menus have the attractive feature of being built on the basis of the intuitive "higher than" relation, which is, arguably, the building block for the empirical study of any given behavioral trait. This section discusses two substantial generalizations of the domain assumption. First, let it be noted that the key implication of an ordered menu is that types selecting the same maximal alternative must form an interval. Hence, we can simply substitute our original domain assumption for the weaker assumption that requires all menus to yield an interval partition of the set of types. Secondly, we show that any menu which does not yield an interval partition can belong to the domain, as long as the domain contains a replica menu; that is, one with a finer interval partition of types than that generated by the menu it replicates.

4.1. **Interval domain.** Claim 1 of Theorem 1 states that, for every pair (x,j), the set of types $\mathcal{T}(x,j)$ is an interval. The reason for this is that, when menu j is ordered, if $t_1 < t_2 < t_3$ are types such that t_1 and t_3 lead to the choice of x while t_2 leads to the choice of $y \neq x$, alternatives x and y would be incomparable for $\triangleleft$. Here, we argue that the strictly weaker property formulated in Claim 1 is sufficient for our purposes. To see that this property constitutes a strict relaxation of the domain of ordered menus, consider three alternatives $\{x,y,z\}$ and the following types: $u_1(x) > u_1(y) > u_1(z)$, $u_2(y) > u_2(z) > u_2(x)$, and $u_3(z) > u_3(x) > u_3(y)$. It is immediate to see that $x \triangleleft z$ and $y \triangleleft z$, but x and y cannot be compared for $\triangleleft$. Hence, menus $\{x,z\}$ and $\{y,z\}$ are ordered, while this is not the case for menus $\{x,y\}$ and $\{x,y,z\}$. Notice, however, that

is strictly larger than its value at t_1, the CDF at type t can be constructed by assigning any value between these two. It can, therefore, be deduced that type t can only be identified in the extreme case in which the values at t_1 and t_2 are equal, leading to $\psi(t) = 0$.

$\{x, y, z\}$ yields the trivial interval structure whereby type 1 chooses x, type 2 chooses y and type 3 chooses z, and hence, this menu can be incorporated into the domain.

Domain of Interval Menus. For every $j \in \mathcal{J}$ and $x \in A_j$, $\mathcal{T}(x, j)$ is an interval.

Corollary 1. *In a domain of interval menus, p satisfies $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity if, and only if, p is a $\mathcal{T}$-RUM.*

Proof of Corollary 1: Claim 1 in the proof of Theorem 1 is now an assumption and can be used to define menu-dependent relations $\lhd_j$ given by: $x \lhd_j y$ if there exist $t \in \mathcal{T}(x, j)$ and $t' \in \mathcal{T}(y, j)$ such that $t < t'$. Notice that $\lhd_j$ is complete on $\{x : \mathcal{T}(x, j) \neq \emptyset\}$ and can replace $\lhd$ in the definition of F. The rest of the proof is analogous and thus, omitted. ■

4.2. **Replica domain.** We now show that the assumption of interval menus can be further generalized to the following domain condition.

Domain of Replica Menus. For every $j \in \mathcal{J}$, there exists $j' \in \mathcal{J}$ such that: (i) j' is an interval menu, and (ii) $t_1, t_2 \in \mathcal{T}(x', j')$ for some x' implies that $t_1, t_2 \in \mathcal{T}(x, j)$ for some x.

Clearly, every domain of interval menus is also a domain of replica menus, but the opposite is not necessarily true. To see this, consider again the same three alternatives $\{x, y, z\}$ and the same types $u_1(x) > u_1(y) > u_1(z)$, $u_2(y) > u_2(z) > u_2(x)$, and $u_3(z) > u_3(x) > u_3(y)$. We claim that $\{x, y, z\}$ is a replica menu of the non-interval menu $\{x, y\}$. We learnt above that $\{x, y, z\}$ is an interval menu, so (i) holds. Moreover, since each alternative in $\{x, y, z\}$ is chosen by a unique type, property (ii) must hold trivially. Therefore, we can include the pair $\{x, y\}$ in the domain as long as menu $\{x, y, z\}$ is also added. Hence, in this particular case, the universal domain is a replica domain despite being neither an interval nor an ordered menu domain. The following result extends Theorem 1 to the case of domains of replica menus.[9]

[9]Intuitively, the result can be extended beyond replica domains, by simply noticing that the replication of menu j could be obtained equivalently by considering a collection of interval menus rather than a unique replica menu.

Corollary 2. *In a domain of replica menus, p satisfies $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity if, and only if, p is a $\mathcal{T}$-RUM.*

Proof of Corollary 2: The proof of Corollary 1 can be reproduced by working exclusively with interval menus, leaving us only to prove that the constructed ψ explains the choices in non-interval menus. The only non-trivial case is $\mathcal{T}(x, j) \neq \emptyset$, where the replica menu j' can be considered. By part (ii) of the definition of replica, we can identify a subset of alternatives $B \subseteq A_{j'}$ such that $\mathcal{T}(x, j) = \bigcup_{y \in B} \mathcal{T}(y, j')$. By $\mathcal{T}$-Monotonicity $p(x, j) = \sum_{y \in B} p(y, j')$. By part (i) of the definition of replica, we know that j' is an interval menu, and thus, all of its choices are rationalized by ψ. Hence, $\sum_{y \in B} p(y, j') = \bigcup_{y \in B} \mathcal{T}(y, j') = \mathcal{T}(x, j)$, which concludes the proof. ■

5. Applications to Decision under Risk

In this section, we discuss the case of decision problems composed of monetary lotteries. We consider the most natural ordered set of types in this setting; namely, a collection of expected utilities, $\{EU_t\}_{t \in \mathcal{T} = \{1,2,\dots,T\}}$, ordered by increasing risk aversion, i.e., by increasing concavity of their respective monetary utilities. Thus, the induced relation $\lhd$ represents the notion of safer lottery. We start by showing, in Proposition 1 below, that some of the most common menus of lotteries used in the literature are ordered. Consequently, whenever the domain is formed by any combination of such menus, $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity lend empirical content to random utility models based on the ordered collection of expected utilities.

We consider three well-known classes of menus. Class (1) encompasses menus in which all lotteries entertain a good-state payoff with some probability p, and a bad-state payoff with probability $1 - p$. This formulation is a generalization of widely used experimental decision problems, such as multiple price lists or convex budget sets.[10] Class (2) involves any menu in which all lotteries belong to the same Marschak-Machina triangle; i.e., they are all composed of the same three monetary payoffs. A

[10]Other influential cases within this class are (i) Tanaka, Camerer and Nguyen (2010) who study a domain in which it is the good-state payoff, and not the probability of the state, that varies across menus, and (ii) the deductible-premium insurance setting of Barseghyan, Molinari and Thirkettle (2019). Also, notice that the limit case, in which the good-state and bad-state payoffs are equal, represents the degenerate lotteries often used, for example, to elicit certainty equivalents.

prominent example of the use of different Marschak-Machina triangles for different menus is Camerer (1989). Class (3) entertains any menu with lotteries offering a possibly different prize with a possibly different probability, and a constant payoff otherwise, such as in the betting example of Chiappori, Salanié, Salanié and Gandhi (2019). We can now establish the following result.

Proposition 1. *Let* $\{EU_t\}_{t\in\mathcal{T}=\{1,2,\ldots,T\}}$ *be ordered by increasing risk aversion. Any domain composed of menus in classes (1), (2) and (3) is ordered.*

Proof of Proposition 1: In all three of the classes considered, we will start by considering a menu j containing two lotteries l_1 and l_2 that are chosen by at least one expected utility type, i.e., $l_1, l_2 \in \{l : \mathcal{T}(l,j) \neq \emptyset\}$. To show that $l_1 \triangleleft l_2$, we simply need to prove that if a type t prefers l_2 to l_1, so does type $s > t$. As a general principle, notice that, whenever both lotteries are chosen by at least one expected utility type, it must be the case that neither lottery first-order stochastically dominates the other.

We then start by considering that menu j belongs to class (1), and denote by q_j the probability of the good state in all the lotteries in the menu. Denote by g_1, g_2 (b_1, b_2) the respective good-state (bad-state) payoffs of the two lotteries at stake. Hence, it must be that $q_j \neq \{0,1\}$ and we can arrange the payoffs, w.l.o.g, as $g_1 > g_2 \geq b_2 > b_1$. Normalize the monetary utilities to $u_t(g_1) = 1$ and $u_t(b_1) = 0$. Let t be one of the types by which lottery l_2 is strictly preferred to lottery l_1. Then $q_j < q_j u_t(g_2) + (1-q_j)u_t(b_2)$, i.e., $\frac{1-u_t(g_2)}{u_t(b_2)} < \frac{1-q_j}{q_j}$. We know that type t is indifferent between receiving b_2 with certainty and receiving g_1 and b_1 with probabilities $u_t(g_2)$ and $1-u_t(b_2)$. Hence, any type $s > t$ is more risk averse than type t and will strictly prefer to receive b_2 with certainty, which implies that $u_s(b_2) > u_t(b_2)$. Similar reasoning proves that $u_s(g_2) > u_t(g_2)$. Hence, $\frac{1-u_s(g_2)}{u_s(b_2)} < \frac{1-u_t(g_2)}{u_t(b_2)} < \frac{1-q_j}{q_j}$, and lottery l_2 must also be preferred to lottery l_1 by the more risk-averse type s. This shows that the menus in class (1) are completely ordered by $\triangleleft$.

We now consider menu j as belonging to class (2). It is evident that lottery l_1 will be preferred to lottery l_2 by type t if, and only if, the slope of the linear indifference curve of the EU type is below the (strictly positive, given the lack of first-order stochastic dominance) slope of the segment connecting these two lotteries. Since types are ordered by increasing slopes, the result follows.

Finally, consider menu j as belonging to class (3), and denote by k_j the constant payoff from all the lotteries in the menu. Further, denote by q_1, q_2, x_1, x_2 the relevant probabilities and prizes of the two lotteries at stake. Lack of first-order stochastic dominance guarantees that one of the lotteries, say l_1, is the dollar-bet, with $x_1 > x_2$, while the other is the probability-bet, with $q_1 < q_2$. By normalization of $u_t(k_j) = 0$, it becomes evident that l_1 is preferred to l_2 if, and only if, $\frac{u_t(x_1)}{u_t(x_2)} \geq \frac{q_2}{q_1}$, which is true if, and only if, the curvature of u_t is sufficiently low. Since types are ordered by increasing curvature, lotteries will be ordered by $\lhd$. This concludes the proof. ■

Proposition 1 shows that the standard experimental and empirical datasets used in the estimation of risk aversion are ordered. Thus, Theorem 1 guarantees that, in these domains, $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity are necessary and sufficient conditions for $\mathcal{T}$-RUMs with expected utilities.

We now show that domains of replica menus, as studied in Section 4, are not particularly taxing in a lottery setting, in the sense that a single Marschak-Machina triangle is sufficient to produce $\lhd$-replicas of any menu of lotteries.

Proposition 2. *Let $\{EU_t\}_{t\in\mathcal{T}=\{1,2,\ldots,T\}}$ be ordered by increasing risk aversion and consider three monetary payoffs $x > y > z$. For any menu of lotteries, there exists a replica menu in the Marschak-Machina triangle defined by these three payoffs.*

Proof of Proposition 2: To see this, consider menu j as inducing the sequence of type intervals $\mathcal{T}(l_1, j) = \{1, 2, \ldots, t_1\}$, $\mathcal{T}(l_2, j) = \{t_1 + 1, t_1 + 2, \ldots, t_2\}$, ..., $\mathcal{T}(l_\kappa, j) = \{t_{\kappa-1} + 1, t_{\kappa-1} + 2, \ldots, T\}$, where the sequence of chosen lotteries may include repetitions. Given the Marschak-Machina triangle, denote by $m(t)$ the (strictly positive) slope of the indifference curves of type t. Notice that $m(t)$ must be strictly increasing in t, because types are ordered by risk aversion. Now, construct the $\lhd$-replica of menu j as follows. Let the vector of probabilities of lottery l_1 be denoted by $q_1 = (\frac{1}{2}, 0, \frac{1}{2})$ and, for $k \in \{2, \ldots, \kappa\}$, let the vector of probabilities of lottery l_k be denoted by $q_k = q_{k-1} + (-\alpha_{k-1}, \alpha_{k-1} + \beta_{k-1}, -\beta_{k-1})$, with $m(t_{k-1}) < \frac{\alpha_{k-1}}{\beta_{k-1}} < m(t_{k-1} + 1)$ and such that $\sum_{k=2}^{\kappa} \alpha_k$ and $\sum_{k=2}^{\kappa} \beta_k$ are both smaller than $\frac{1}{2}$, thus guaranteeing that all vectors correspond to lotteries within the triangle. It is then evident that lottery l_k is maximal for, and only for, types $\{t_{k-1} + 1, \ldots, t_k\}$, as desired. ■

The proof of Proposition 2 constructs elementary menus which exactly reproduce the choice intervals of any arbitrary, not-necessarily ordered, menu of lotteries. Thus, any domain of menus of lotteries can be extended into a replica domain simply by adding menus of lotteries from a given Marschak-Machina triangle. As an immediate consequence of Corollary 2, and for any such extended domain, $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity are necessary and sufficient conditions for $\mathcal{T}$-RUMs with expected utilities.

6. Extensions

In this section, we discuss three theoretical extensions of the model that prove important in many applications. First, we extend the setting to incorporate a simple trembling stage in order to accommodate the choice of dominated alternatives. Next, we contemplate the analysis of different subpopulations, and study the case in which their behaviors are expected to be related by first-order stochastic dominance. Finally, we briefly discuss how to deal with infinite type spaces and menus.

6.1. **Tremble.** $\mathcal{T}$-RUMs cannot explain the choice of alternatives that are not maximal for any type in $\mathcal{T}$, such as first-order stochastically dominated lotteries under expected utility types. Sometimes, however, the non-negligible choice of non-maximal alternatives can occur, in which case the analyst may wish to extend the model to accommodate this behavioral regularity. A convenient way to do this is to include trembling behavior. Given a menu j, denote by D_j the set of alternatives that are dominated by all utility types, i.e., $D_j = \{x \in A_j : \mathcal{T}(x, j) = \emptyset\}$. The $\mathcal{T}$-RUM with tremble ($\mathcal{T}$-RUMT) is then defined by means of a probability distribution ψ and a tremble function $\lambda : \mathcal{J} \to [0, 1]$, with behavior given by: (i) whenever $D_j = \emptyset$, choices are determined according to ψ and (ii) whenever $D_j \neq \emptyset$, choices from $A_j \setminus D_j$ are determined according to ψ with probability $(1 - \lambda_j)$; and alternatives from D_j are chosen otherwise. That is, the behavior over the set of maximal alternatives is governed by the $\mathcal{T}$-RUM ψ but, whenever non-maximal alternatives exist, a tremble probability λ_j leads to the mentioned inconsistent choices.[11] $\mathcal{T}$-RUMT is a simple model, and its characterization follows immediately from the analysis in Theorem 1. Given the

[11]For ease of presentation, we have formulated a class of tremble models where the tremble distribution across non-maximal alternatives is left unstructured. The analyst may focus on maximal alternatives exclusively or select the tremble model that is theoretically more sound, appears more

stochastic choice function p, denote by $\bar{p}$ the conditional stochastic choice function:[12]

$$\bar{p}(x,j) = \begin{cases} \frac{p(x,j)}{1-\sum_{x\in D_j} p(x,j)}, & \text{if } x \in A_j \setminus D_j; \\ 0, & \text{otherwise.} \end{cases}$$

We then have:

Corollary 3. *In a domain of ordered menus, p is a $\mathcal{T}$-RUMT if, and only if, $\bar{p}$ satisfies $\mathcal{T}$-Monotonicity.*

Proof of Corollary 3: Necessity is immediate. To see sufficiency, notice that $\bar{p}$ must satisfy $\mathcal{T}$-Extremeness by construction and, hence, $\bar{p}$ must be a $\mathcal{T}$-RUM. The techniques in the proof of Theorem 1 can be used to construct the corresponding distribution ψ, and the definition of $\lambda_j = \sum_{x\in D_j} p(x,j)$ completes the $\mathcal{T}$-RUMT. The claim then follows immediately. ■

6.2. **Subpopulations.** In many applications, the analyst envisions a model in which different subpopulations are expected to behave differently, and wishes to establish a relationship across the behaviors of these subpopulations. We illustrate this idea with the intuitive case in which an ordered characteristic (say age or income, or a gender dummy or one of a set of treatment dummies) is such that higher subpopulations have type distributions that are first-order stochastically dominating, i.e., they are skewed towards higher types. Formally, let $\mathcal{G} = \{1, \dots, G\}$ be a partition of the population into G subpopulations. A monotone-in-characteristics $\mathcal{T}$-RUM is defined by a collection of distributions $\{\psi_g\}_{g\in\mathcal{G}}$, one for each subpopulation, satisfying the property that, for every $g, g' \in \mathcal{G}$ with $g < g'$ and every $t \in \mathcal{T}$, $\sum_{s=1}^{t} \psi_g(s) \geq \sum_{s=1}^{t} \psi_{g'}(s)$. Denoting by p_g the stochastic choice function of subpopulation g, we then have:

Corollary 4. *In a domain of ordered menus, $\{p_g\}_{g\in\mathcal{G}}$ is a monotone-in-characteristics $\mathcal{T}$-RUM if, and only if, every p_g satisfies $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity and, for every $g < g'$ and (x,j), it is the case that $\sum_{y\in A_j : y \trianglelefteq x} p_g(y,j) \geq \sum_{y\in A_j : y \trianglelefteq x} p_{g'}(y,j)$.*

realistic, or proves easier to implement. In Section 8, we use the simplest tremble model within the class, where λ is menu-independent and choices over non-maximal alternatives are uniformly random.

[12]When $\sum_{x\in D_j} p(x,j) = 1$, define the (irrelevant) conditional probability as $\bar{p}(x,j) = \frac{|\mathcal{T}(x,j)|}{|\mathcal{T}|}$.

Proof of Corollary 4: Necessity is immediate. To see sufficiency, we can construct ψ_g using Theorem 1 on p_g. Since $\sum_{y\in A_j:y\trianglelefteq x} p_g(y,j)$ defines the CDF of group g at type $\max \mathcal{T}(x,j)$, it is evident that $\sum_{s=1}^{t}\psi_g(s) \geq \sum_{s=1}^{t}\psi_{g'}(s)$ must hold for every $t \in \mathcal{T}^I$. For this dominance to hold at every type, it is necessary only to guarantee that ψ_g for types that cannot be identified is constructed consistently across groups. This can be always done, e.g., by fixing $\psi_g(t) = 0$ whenever $t \notin \mathcal{T}^I$, or by dividing the mass uniformly over the non-identified types. This concludes the proof. ■

6.3. **Infinite types and menus.** Some settings make use of an infinite type space, with menus defined on a real space. Our analysis can be easily extended to this setting, while applying standard continuity and measurability assumptions. Here, $\mathcal{T}$ is a subset in the reals, denoting the infinite type space. We can consider a continuous density ψ over $\mathcal{T}$, describing the mass with which types are realized. Given the $\mathcal{T}$-RUM with density ψ, the probability of choice of any (measurable) subset of alternatives B in menu j is $\psi(\bigcup_{x\in B}\mathcal{T}(x,j))$.

With respect to the characterization, we need to reformulate the properties, $\mathcal{T}$-Extremeness and $\mathcal{T}$-Monotonicity, over measurable sets. The first of these properties requires that only subsets of alternatives that are maximal for a non-zero measurable set of types can have strictly positive mass, while the other operates over subsets of alternatives and menus that produce measurable sets of types. The proof of the characterization result follows the same steps as in Theorem 1. The infinite $\mathcal{T}$-RUM can be recovered from data by assigning mass to the half-open intervals of types of the form $(-\infty, \sup \mathcal{T}(x,j))$. The mass of these types is, as in the finite version, the cumulated probability of choice of all alternatives in menu j that are lower than x.

7. ϵ-rationalizability and Strongly Consistent Estimators

We now return to our base model described in Section 3, and present several results relating to the estimation of $\mathcal{T}$-RUMs when finite data are available. Formally, the data form a map $z : X \times \mathcal{J} \to \mathbb{Z}_+$, describing the number of observed instances in which each alternative is chosen in each menu, with $z(x,j) > 0$ implying that $x \in A_j$. For every $j \in \mathcal{J}$, we denote by $z(\cdot,j)$ the vector describing the observed choices in menu j, and by $Z_j = \sum_{x \in A_j} z(x,j) > 0$ the total number of observations for this menu. Observed choice frequencies in menu j are therefore $\frac{z(\cdot,j)}{Z_j}$.

Notice first that, when data is generated by a $\mathcal{T}$-RUM, observed choice frequencies always satisfy $\mathcal{T}$-Extremeness, i.e., $\frac{z(\cdot,j)}{Z_j} > 0$ implies that $\mathcal{T}(x,j) \neq \emptyset$. However, $\mathcal{T}$-Monotonicity may be violated due to sampling issues. Intuitively, a good estimator of the underlying distribution of types should deviate minimally from observed choice frequencies, up to the point at which $\mathcal{T}$-Monotonicity holds, which is when a reasonably good distribution $\widehat{\psi}$ is constructed.

We formalize this idea by providing a generalization of the rationalizability notion embedded in $\mathcal{T}$-RUMs, which allows for menu-dependent perturbations of the underlying distribution of types. Thus, in order to accommodate violations of $\mathcal{T}$-Monotonicity, we consider the case in which different probability distributions over $\mathcal{T}$ rationalize choices in different menus. We then evaluate the deviation of these menu-dependent probability distributions from a fundamental distribution of types. Formally, given $d : \Psi \times \Psi \to \mathbb{R}_+$ and $f : \mathbb{R}^J_+ \to \mathbb{R}_+$, we say that the data are ϵ-rationalizable if ϵ is the smallest value for which there exists one probability distribution ψ, and distributions $\{\psi_j\}_{j \in \mathcal{J}}$, with ψ_j rationalizing the data in menu j, such that $f(d(\psi,\psi_1),\ldots,d(\psi,\psi_J)) = \epsilon$. That is, d captures the distance (or divergence) between the fundamental distribution of type ψ with the one explaining menu j, ψ_j, and f aggregates all such deviations across menus. Technically, we assume that d: (i) is continuous, (ii) strictly increases if all the non-zero differences between the masses of types increase, and (iii) takes the value 0 if, and only if, the two distributions coincide. In the standard maximum-likelihood or least squares approaches, for instance, d is a map that operates on each subset of types $\mathcal{T}(x,j)$, considering the logarithmic ratio or the square of the difference between the two distributions, and additively aggregating them all. Alternatively, in the separation technique in Apesteguia and Ballester (2020), d corresponds to one such subset of types $\mathcal{T}(x,j)$, namely, the one with the largest observed ratio-deviation across distributions. The map f is assumed: (i) to be continuous, (ii) to be strictly increasing when all non-zero components increase and, (iii) to take the value 0 at, and only at, its origin. Thus, for example, f may be an additive operator, as in maximum-likelihood or least squares estimations, or a maximum operator, as in the separation technique used in Apesteguia and Ballester (2020). Thus we have:

Corollary 5. *z is ϵ-rationalizable for some value of ϵ. Moreover, z is 0-rationalizable if, and only if, the stochastic choice function defined by its choice frequencies is a $\mathcal{T}$-RUM.*

Proof of Corollary 5: We start by proving the first part. Consider any menu $j \in \mathcal{J}$ in isolation. The observed choice frequencies in this menu satisfy $\mathcal{T}$-Extremeness and, trivially, $\mathcal{T}$-Monotonicity. Hence, the set of distributions rationalizing the data in menu j, which we denote by Ψ_j, must be non-empty. Furthermore, given finiteness of menus, types and domain, $\Psi \times \Psi_1 \times \cdots \times \Psi_J$ is a compact set. Hence, given the continuity of d and f, the result follows. To prove the second part, notice that the properties of d and f make 0-rationalizability equivalent to the existence of a distribution ψ that belongs to $\cap_j \Psi_j$, thus concluding the proof. ■

The first part of the result states that, with ϵ large enough, any finite data generated by a $\mathcal{T}$-RUM can be explained by allowing sufficiently large menu-dependent perturbations.[13] The continuity of maps d and f guarantees that all data can be assigned a minimal rationalizability value. The second part of the result describes how ϵ-rationalizability constitutes a generalization of rationalizability by a $\mathcal{T}$-RUM, as the latter requires no perturbation whatsoever, i.e., $\epsilon = 0$.

Hence, when finite data violate $\mathcal{T}$-Monotonicity, a natural goodness-of fit measure for the model is given by the (smallest, by definition) magnitude of ϵ that yields ϵ-rationalizability. Moreover, the distribution $\widehat{\psi}$ on Ψ that yields that (minimal) value ϵ represents an intuitive extremum estimator. We now show that this class of estimators is strongly consistent.[14]

Theorem 2. *$\widehat{\psi}$ is strongly consistent.*

Proof of Theorem 2: Denote by $\mathcal{SCF}$ the set of all stochastic choice functions. Consider the map $g : \Psi \times \mathcal{SCF} \to \mathbb{R}_+$ defined by $g(\psi, p) = f(d(\psi, \Psi_1^p), \ldots, d(\psi, \Psi_J^p))$, where

[13]It follows immediately from the notion of ϵ-rationalizability that any stochastic choice function satisfying $\mathcal{T}$-Extremeness can be rationalized in this way. Notice that, if $\mathcal{T}$-Extremeness does not hold, it may be more appropriate to work with the $\mathcal{T}$-RUMT version of the model; the ϵ-rationalizability concept can be modified accordingly and, then, all stochastic choice functions can be ϵ-rationalized for some value of ϵ.

[14]Naturally, over the set of identifiable types.

Ψ_j^p is the set of probability distributions rationalizing the probabilities of choice defined by p over menu j. Notice that g is an extension of the map that is used in the notion of ϵ-rationalizability to the space of all stochastic choice functions, not only those that may be yielded by finite $\mathcal{T}$-RUM experiments. Notice also that, given the properties of d and f, (i) map g has zero value if, and only if, $\psi \in \cap_j \Psi_j^p$, (ii) is continuous and (iii) is strictly increasing whenever all non-zero components increase. Consider a sequence of data functions $\{z^n\}_{n=1}^{\infty}$ yielded by $\mathcal{T}$-RUM experiments, with $\lim_{n\to\infty} Z_j^n = \infty$ for every $j \in \mathcal{J}$. Evidently, the data z^n implicitly define a stochastic choice function, which we denote by z^n to avoid further notation, by considering the frequencies of choice over each menu. The data estimator for z^n is $\widehat{\psi}^n = \arg\min_{\psi\in\Psi} g(\psi, z^n)$; in other words, the distribution of types that minimizes the continuous loss function g with respect to data z^n.

Next, suppose that the data are generated by a $\mathcal{T}$-RUM with probability distribution $\psi^* \in \Psi$. Consider menu j. For every alternative $x \in A_j$ such that $\psi^*(\mathcal{T}(x,j)) = 0$, either because $\mathcal{T}(x,j) = \emptyset$ or because no mass is associated to the types for which x is maximal, we know that $\frac{z^n(x,j)}{Z_j^n} = 0$ always holds. For every alternative $x \in A_j$ such that $\psi^*(\mathcal{T}(x,j)) > 0$, the multinomial i.i.d. nature of choices in the menu guarantees that $\frac{z^n(x,j)}{Z_j^n}$ converges, almost surely, to $\psi^*(\mathcal{T}(x,j))$. Thus, the finiteness of each menu and of $\mathcal{J}$ guarantees that the sequence of stochastic choice functions z^n must converge, almost surely, to the stochastic choice function generated by ψ^*. It then follows immediately that $\widehat{\psi}^n$ converges, almost surely, to ψ^*. This concludes the proof. ■

Theorem 2 shows that the estimators $\widehat{\psi}$ are strongly consistent. The proof is based on the fact that the finiteness of the domain guarantees that the stochastic choice function defined by choice frequencies must converge, almost surely, to the stochastic choice function determined by the true distribution ψ^*. The result follows from the technical properties of d and f and the fact that the estimator can be seen as the minimization of divergence between a distribution and a stochastic choice function.

8. Empirical Illustration

We now illustrate our framework and results using the experimental dataset analyzed in Apesteguia and Ballester (2020), in which 87 UCL undergraduates choose lotteries from menus of sizes 2, 3, and 5, from the nine equiprobable monetary lotteries described in Table 1. Each participant faces a total of 108 different menus of lotteries, including all 36 binary menus; 36 menus with 3 alternatives, out of the possible 84; and another 36 menus with 5 alternatives, out of the possible 126. Random individual processes, without replacement, were used for the selection and order of presentation of the menus of 3 and 5 alternatives, and for the location of the lotteries on the screen and the monetary prizes within a lottery. There were two treatments, NTL and TL. Treatment NTL was a standard implementation, that is, with no time constraint. In treatment TL, subjects had to select a lottery within 5, 7 and 9 seconds from the menus with 2, 3, and 5 alternatives, respectively.[15]

Table 1. Lotteries

$l_1 = (17)$	$l_4 = (30, 10)$	$l_7 = (40, 12, 5)$
$l_2 = (50, 0)$	$l_5 = (20, 15)$	$l_8 = (30, 12, 10)$
$l_3 = (40, 5)$	$l_6 = (50, 12, 0)$	$l_9 = (20, 12, 15)$

We use expected utility and CRRA, i.e., monetary utilities $\frac{x^{1-\omega}}{1-\omega}$, whenever $\omega \neq 1$, and $\log x$ for $\omega = 1$, where ω represents the risk-aversion coefficient.[16] Given the decision problems considered in the experiment, this gives exactly 30 relevant types, which are ordered by increasing risk aversion, as described in Columns 1-3 of Table 2. Column 1 reports the type number, Column 2 reports the upper bound of ω corresponding to that type, and Column 3 reports the preferred lottery of the respective type. We describe the preference of the first type and then specify the pair(s) of alternatives that flip from the previous type. Given the types, it is easy to see that all menus of lotteries are ordered by $\lhd$ and, hence, our basic framework is applicable.[17]

There are several menus for which some alternatives are not chosen by any type, such as, for example, the binary menu $\{l_5, l_9\}$, where l_5 first-order stochastically dominates l_9. The observation of non-negligible choice probabilities for some of these alternatives advises use of the tremble version of the model, $\mathcal{T}$-RUMT. The model is operationalized in the simplest possible form, using a constant tremble parameter across all choice

[15]Experimental payoffs were determined by randomly selecting one menu, and awarding subjects according to their choice from that menu.

[16]Since lotteries l_2 and l_6 involve 0 payoffs, we assume a small fixed positive background consumption.

[17]Note that some of these menus do not fall within the classes covered in Proposition 1, thus reinforcing our claim regarding the wide applicability of the setting studied in this paper.

Table 2. Preferences and Estimation Results

Preferences			Estimated $\mathcal{T}$-RUMTs			
ID	ω		All Obs	TL	NTL	Only Binary
1	-4.148	2-6-3-7-4-8-5-9-1	0.1166	0.1173	0.1158	0.2197
2	-0.518	$(1,9)$	0	0	0	0
3	-0.313	$(3,6)$	0.1009	0.0937	0.1081	0.0673
4	-0.083	$(4,7)$	0	0	0	0
5	0.065	$(5,8)$	0	0	0	0
6	0.154	$(4,6)$	0	0	0	0
7	0.209	$(1,8)$	0	0	0	0
8	0.229	$(3,2)$ and $(7,6)$	0.0561	0.0713	0.0289	0
9	0.258	$(5,6)$	0	0	0.031	0
10	0.262	$(1,6)$	0.0274	0.009	0.0263	0
11	0.273	$(5,7)$	0	0	0	0
12	0.339	$(4,2)$ and $(8,6)$	0	0	0	0
13	0.342	$(1,7)$	0	0	0	0
14	0.358	$(5,2)$ and $(9,6)$	0	0	0	0
15	0.363	$(1,2)$	0	0	0	0
16	0.374	$(7,2)$	0	0	0	0
17	0.408	$(8,2)$	0	0	0	0
18	0.443	$(9,2)$	0.0915	0.067	0.1172	0.1545
19	0.516	$(4,3)$ and $(8,7)$	0.0079	0.013	0	0
20	0.607	$(6,2)$	0	0	0	0
21	0.652	$(5,3)$ and $(9,7)$	0	0	0	0
22	0.808	$(1,3)$	0	0	0	0
23	0.844	$(8,3)$	0	0	0.009	0
24	1.000	$(9,3)$	0.1388	0.1482	0.1182	0.0645
25	1.124	$(5,4)$ and $(9,8)$	0	0	0	0
26	1.309	$(1,4)$	0	0	0	0
27	2.000	$(7,3)$	0.0497	0.0473	0.0594	0
28	2.826	$(9,4)$	0.0489	0.025	0.0632	0
29	4.710	$(1,5)$	0	0	0.0373	0.0573
30	∞	$(8,4)$	0.3622	0.4082	0.2856	0.4368
			Estimated λ			
			0.2516	0.2829	0.2199	0.1744
			Log Likelihood			
			-10.7168	-5.3898	-5.2938	-27.1106

problems, and uniform selection of non-maximal alternatives within each menu. The ML data estimator is then analyzed.[18] Table 2 reports the estimated densities of the $\mathcal{T}$-RUMT using all the data (Column 4); the treated data (Column 5); the non-treated data (Column 6); and the binary data, that is, the aggregated treated and non-treated data (Column 7).

[18]The data and estimation programs are available for use on our websites.

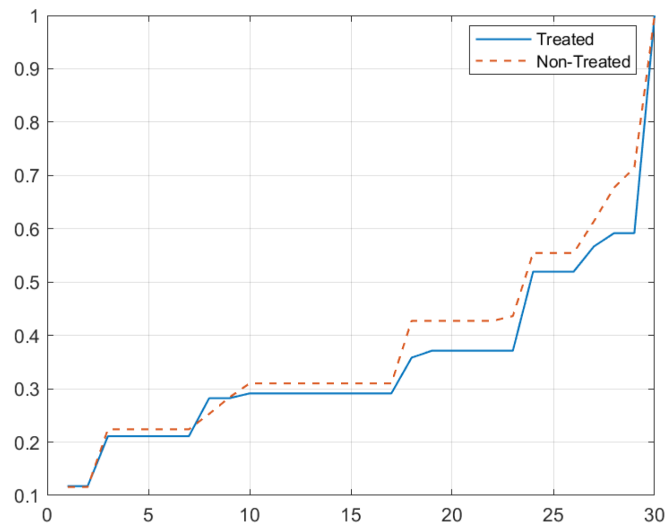

FIGURE 1.—CDFs of the estimated RTMTs of the treated and non-treated data.

The estimated $\mathcal{T}$-RUMT using all the data shows that a relatively small number of types, a third of the total, is sufficient to capture the behavior of the population in this experiment. The results show a high degree of heterogeneity including a very significant proportion of highly risk-averse types. The percentage of types with close to or higher than logarithmic curvature (type 24 and above) is 60%, with more than a third of all decisions belonging to the highest type, type 30, which exhibits risk-aversion levels as high as 4.7 and above. Interestingly, the results also show that a relevant proportion of the decisions reflect highly risk-seeking attitudes (22%), and even extreme risk-seeking, (12% of all decisions correspond to the lowest type, type 1, that is, risk-aversion coefficients below -4.8). The estimated probability of tremble is .25, thus confirming the behavioral relevance of non-maximal alternatives.

Comparison of the estimated $\mathcal{T}$-RUMTs in the treatment-control data yields some interesting findings. See Columns 5 and 6 in Table 2 for the densities, and Figure 1 for a representation of the corresponding CDFs. Both $\mathcal{T}$-RUMTs allocate masses to similar preferences, although there is a slight tendency towards more risk aversion in the treated data. The masses allocated to types with close to or higher than logarithmic curvature are 63% and 56% in the treated and non-treated data, respectively. This can also be seen in Figure 1, where, starting from the 10th preference, the CDF of the treated $\mathcal{T}$-RUMT dominates that of the non-treated. Note, in addition, that the estimated tremble with the treated data is about 22% higher than with the non-treated data. In sum, there are differences between treatment and control, suggesting a behavioral shift towards higher risk aversion, and towards greater choice inconsistencies, although the changes do not appear dramatic. This may be due to the fact that the subject population was highly risk averse to begin with, thus leaving little room for the identification of stronger effects in the treatment-control comparison, and also to the fact that the imposed time constraints were probably not strict enough. Finally, Column 7 reports the estimated $\mathcal{T}$-RUMT using all the binary data, which, interestingly, reveal risk-aversion levels analogous to those obtained previously.

References

[1] Andersson, O., H.J. Holm, J.R. Tyran and E. Wengström (2018). "Robust Inference in Risk Elicitation Tasks," mimeo.

[2] Apesteguia, J. and M.A. Ballester (2020). "Separating Predicted Randomness from Residual Behavior," *Journal of the European Economic Association*, forthcoming.

[3] Apesteguia, J., M.A. Ballester and A. Gutierrez (2020). "Random Models for the Joint Treatment of Risk and Time Preferences," mimeo.

[4] Apesteguia, J., M.A. Ballester and J. Lu (2017). "Single-Crossing Random Utility Models," *Econometrica*, 85(2):661–674.

[5] Barseghyan, L., F. Molinari and M. Thirkettle (2019). "Discrete Choice under Risk with Limited Consideration," mimeo.

[6] Barsky, R.B., F.T. Juster, M.S. Kimball and M.D. Shapiro (1997). "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 112(2):537–79.

[7] Camerer, C., (1989). "An Experimental Test of Several Generalized Utility Theories", *Journal of Risk and Uncertainty*, 2: 61-104

[8] Cattaneo, M., X. Ma, Y. Masatlioglu and E. Suleymanov (2020). "A Random Attention Model," *Journal of Political Economy*, forthcoming.

[9] Chiappori, P-A, B. Salanié, F. Salanié and A. Gandhi (2019). "From Aggregate Betting Data to Individual Risk Preferences," *Econometrica*, 87(1):1–36.

[10] Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, 97(5):1921–1938.

[11] Cohen, A. and L. Einav (2007). "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97(3):745–788.

[12] Coller, M. and M.B. Williams (1999). "Eliciting Individual Discount Rates," *Experimental Economics*, 2(2):107–27.

[13] Cunha, F., J.J. Heckman and S. Navarro (2007). "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds," *International Economic Review*, 48(4):1273–1309.

[14] Dardanoni, V., P. Manzini, M. Mariotti and C.J. Tyson (2020). "Inferring Cognitive Heterogeneity from Aggregate Choices," *Econometrica*, forthcoming.

[15] Filiz-Ozbay, E. and Y. Masatlioglu (2020). "Progressive Random Choice," mimeo.

[16] Greene, W.H. and D.A. Hensher (2010). *Modeling Ordered Choices*, Cambridge University Press.

[17] Gul, F. and W. Pesendorfer (2006). "Random Expected Utility," *Econometrica*, 74(1):121–146.

[18] Holt, C.A. and S.K. Laury (2002). "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5):1644–1655.

[19] Jagelka, T. (2020). "Are Economists' Preferences Psychologists' Personality Traits? A Structural Approach," mimeo.

[20] Mirrlees, J.A. (1971). "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2):175–208.

[21] Spence, A.M. (1974). *Market Signaling: Information Transfer in Hiring and Related Screening Processes*, Harvard University Press.

[22] Tanaka, T., Camerer, C. and Q. Nguyen (2010) "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review*, 100:1, 557571.

[23] Warner, J.T. and S. Pleeter (2001). "The Personal Discount Rate: Evidence from Military Downsizing Programs," *American Economic Review*, 91(1):33–53.