

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 understanding a variety of economic behaviors. This paper studies the micro-foundations of ordered random utility models with the objective of meeting empirical requirements. This is done by working with arbitrary collections of ordered menus of alternatives, and by making no parametric assumptions about the type distribution. The model is characterized by a simple monotonicity axiom. Goodness-of-fit measures are proposed, with proof provided of the strong consistency of extremum estimators defined upon them. A statistical test for the model is also provided.

Keywords: Random utility model; Ordered utilities; Arbitrary domains.

JEL classification numbers: C00; D00

Date: May, 2023.

*We thank Jay Lu and Christoph Semken for helpful comments and Angelo Gutierrez for outstanding research assistance. Financial support by FEDER/Ministerio de Ciencia de Innovación (Agencia Estatal de Investigación) through Grant PID2021-125538NB-I00 and through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), and Balliol College is gratefully acknowledged.

[†]ICREA, Universitat Pompeu Fabra and BSE. E-mail: jose.apestegui@upf.edu.

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

1. INTRODUCTION

Settings in which a collection of ordered utilities describes a specific behavioral trait, with higher utilities selecting higher alternatives, are common in economics. For example, consider the case of decisions under risk; utilities can be ordered by risk aversion, with higher utilities leading to safer lotteries. In such settings, heterogeneity is often modeled as a probability distribution over the set of utilities, i.e., as an ordered random utility model. Recently, Apesteguia, Ballester and Lu (2017) introduce ordered models in the standard theoretical setting of stochastic choice. They derive a set of properties that characterize the single-crossing random utility model (SCRUM), in which all alternatives and utilities can be ordered (à la single-crossing), and a mixture over the utilities rationalizes the stochastic choices in all possible menus of alternatives.

Bringing theoretical models to empirical applications is sometimes challenging. The practitioner often faces a small number of menus of alternatives, instead of the standard theoretical assumption of having choice data for every menu. This immediately highlights the need for properties that are more suitable for restricted domains of menus. Furthermore, in the present setting with ordered models, some families of utilities may not universally produce ordered choices, but they may do so in the restricted domain of observed menus. In this paper, we study theoretically ordered random utility models for arbitrary domains of menus.

We consider the situation where an analyst has a dataset describing the observed choice frequencies over an arbitrary collection of menus of alternatives. The dataset represents either the aggregate choices of a heterogeneous population or the repeated choices of an individual subject to intrapersonal variation. The analyst has in mind a specific collection of ordered utilities \mathcal{T} , maybe due to its analytical convenience or empirical prominence in the literature, such as CRRA expected utilities in the treatment of risk. The analyst then wonders whether a distribution over such ordered collection of utilities accounts for the data, that is, whether there is a random utility model over \mathcal{T} , which we refer to as \mathcal{T} -RUM, that can rationalize the data.

The first contribution of this paper is to provide a characterization of the choice frequencies that can be generated by a \mathcal{T} -RUM. We use the following property, that we call \mathcal{T} -Monotonicity. Suppose that the types that lead to alternatives B_1 in menu A_1 are a subset of the types resulting in the choice of alternatives B_2 in menu A_2 . In this case, \mathcal{T} -Monotonicity states that the cumulative choice frequency of alternatives in B_1 within menu A_1 must be smaller than the cumulative choice frequency of alternatives

in B_2 within menu A_2 . Interestingly, we show that, when the menus are ordered, \mathcal{T} -Monotonicity is both necessary and sufficient.

The proof of the characterization theorem is fully constructive, and hence, when the property is satisfied by the data, we can determine the underlying type distribution that explains all choices. Moreover, we show that the model is uniquely identified on a subset of the set of types that we characterize. To ensure the applicability of our findings, we present a simple linear algorithm for the analysis of \mathcal{T} -Monotonicity. We then discuss a generalization of the ordered domain assumption; we note that our analysis goes through even if the alternatives within a menu are not ordered, provided that every alternative in every menu is chosen by an interval of types.

The first part of the paper concludes with an extension of the model, where we allow for the possibility of observing choices of dominated alternatives, that is, alternatives that are never maximal and are hence predicted to have zero choice probabilities in the \mathcal{T} -RUM. Our approach is to minimally extend our main model, by incorporating the possibility of mistakes in decision-making.

Given our interest in connecting theory with empirics, we then present some results of econometric interest dealing with finite data. First, observed choice frequencies, due to sampling issues, may violate \mathcal{T} -Monotonicity and, consequently, we introduce choice-based goodness-of-fit measures. These are based on perturbing the underlying distribution of types. Our approach considers the smallest perturbation necessary to account for all observed choice frequencies, and implicitly defines a class of extremum estimators. Most importantly, subsequent analysis shows that any estimator within this class is strongly consistent. That is, as the number of observations per choice problem increases, the estimator converges to the true distribution of probabilities over the types. Second, we show how the model can be statistically tested. We exploit the i.i.d. nature of \mathcal{T} -RUM which enables us to interpret the model as a collection of independent multinomial distributions with parameters linked through the distribution of types. We then propose an aggregated Pearson statistic to statistically test the model.

In the Online Appendix we provide a detailed and exhaustive guide for implementing the model, and illustrate each step in our guide using an existing experimental dataset that involves decision problems over lotteries.

2. RELATED LITERATURE

We start by elaborating on the main differences between this paper and Apesteguia, Ballester and Lu (2017).¹ As mentioned in the introduction, there are two fundamental differences between the two settings: (1) in this paper we use an arbitrary domain of menus which allows us to work with any dataset and with any collection of utilities that produce ordered choices, and (2) instead of deriving the collection of utilities from observed choices, we consider a fixed collection of utilities. The first point helps us bring the model to data, at the cost described in the second point.² An immediate consequence of these differences is that the characterizing properties of SCRUM are useless in the present setting. This is the case because these properties, formulated over the universal domain, apply to pairs of menus related by set inclusion. With arbitrary domains, data may comprise any collection of menus of alternatives, such as disjoint menus, and the properties would be emptyly satisfied. Conversely, our property of \mathcal{T} -Monotonicity contains insights of the properties of Regularity and Centrality in the treatment of SCRUMs. However, \mathcal{T} -Monotonicity is only applicable insofar as we are using a specific collection of utilities \mathcal{T} and hence it is not applicable to the setting of SCRUMs.

In a recent theoretical contribution, Filiz-Ozbay and Masatlioglu (2023) 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, ordered 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 exploiting the single-crossing condition. In a risk environment, Barseghyan, Molinari and Thirkettle (2021) combine a family of expected utilities ordered à la SCRUM with models of limited attention. They show that with sufficiently rich variation in the menus of lotteries, point-identification of the parameters for risk and attention can be achieved, and provide a simple method to

¹See also Petri (2023), Turansick (2022) and Valkanova (2021) for further theoretical results related to SCRUM.

²The stochastic choice theory literature is turning to the issue of 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, while Cattaneo, Ma, Masatlioglu and Suleymanov (2020) establish non-parametric identification results for their random attention model.

compute a likelihood-based estimator. Our paper contributes by providing foundational properties to the ordered models incorporated in this application, and by showing that a large class of consistent estimators, including maximum-likelihood, are readily available to the practitioner even when data is scarce. Our theoretical framework sets the basis for the consideration of other behavioral variables, such as attention, when data are scarce. For example, one could start by assuming a given ordered family of utilities and a given attention model and then reformulate the property of \mathcal{T} -Monotonicity to account for attention considerations.

Chiappori, Salanié, Salanié and Gandhi (2019) also impose the single-crossing condition on individual risk preferences in a parimutuel horse-racing setting. The authors first derive the equilibrium of the model and give necessary conditions on the data implied at equilibrium and, when the data is rich, they ultimately identify the model. Our analysis shares with this paper the idea of using the critical type that determines the jump from one choice to the next one in the order. We show that this logic can be formalized as a property, \mathcal{T} -Monotonicity, that it is not only necessary but also sufficient for ordered-choice rationalizability. Our results then pave the way for characterizing stochastic data that emerge from equilibrium conditions.

A series of applied papers have implemented parametric versions of the random utility model, over an ordered collection of utilities, to estimate a specific behavioral trait, most frequently, risk aversion.³ Barsky, Juster, Kimball and Shapiro (1997) is one of the first examples of the use of this methodology, where the ordered structure of a menu 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 menu of deductibles and premiums, thereby facilitating the estimation of risk aversion. Andersson, Holm, Tyran and Wengström (2020) use menus 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 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

³See Collier and Williams (1999) and Warner and Pleeter (2001) for similar estimation exercises within the context of time preferences, or Apesteguia, Ballester and Gutierrez (2023) and Jagelka (2023) for joint estimations of risk and time preferences.

references cited therein, and Greene and Hensher (2010) for a survey). The papers in this literature focus on identification relying on the existence of a relationship between the probability of choice of any alternative and the mass of types for which the alternative is optimal, which, given the structure, takes the form of an interval. However, there are no axiomatic exercises of any kind in this literature. Hence, the novelty of our paper is to bring the ordered choice logic from the applied and econometrics literature to a revealed preference setting that imposes minimal requirements on the data structure, to provide a novel and easily testable property, \mathcal{T} -Monotonicity, and to show that it fully axiomatizes the ordered random type model.

3. ORDERED RANDOM UTILITY MODELS AND ORDERED DOMAINS

Let X be the set of all alternatives. We fix an ordered collection of utilities $\{U_t\}_{t \in \mathcal{T}}$, where $\mathcal{T} = \{1, 2, \dots, T\}$.⁴ Given the ordinal nature of all our results, 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. In each menu of alternatives, one of the utility functions is independently realized according to ψ and maximized, thus determining the choice. Menus are finite subsets of alternatives. We work with an arbitrary collection of menus, $\{A_j\}_{j \in \mathcal{J}}$, where $\mathcal{J} = \{1, 2, \dots, J\}$. Given the \mathcal{T} -RUM with distribution ψ , the probability of choosing 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 $x_l \triangleleft x_h$, and, as usual, $x \trianglelefteq y$ whenever $x \triangleleft y$ 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 or altruism, and hence the notion of a higher alternative corresponds either to the notion of a safer lottery or to that of a more altruistic distribution. We now introduce the only relevant assumption in the

⁴For purposes of exposition, we assume no role for indifferences, i.e., utility functions are strict over the domain for which data exist.

paper: i.e., that every menu in the domain is ordered, in the sense that its maximal alternatives can be ordered by \preceq .

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

Domains composed of ordered menus appear naturally when studying a particular behavioral trait. The following three economic applications illustrate this point:

- (1) **Expected Utility.** Let $\{EU_t\}_{t \in \mathcal{T}=\{1,2,\dots,T\}}$ be a collection of expected utilities ordered by increasing risk aversion, i.e., by increasing concavity of their respective monetary utilities. Thus, in this case, the induced relation \triangleleft represents the notion of a safer lottery. Classical results on second order stochastic dominance, such as Hammond (1974), guarantee that standard domains of menus of lotteries used in the study of risk aversion are ordered.⁵
- (2) **Quasi-Linear Utility.** Consider pairs of the form (q, w) , with q in an ordered set Q and w representing money. For example, q may describe the quality of a product, and w the income after salary.⁶ Consider a collection of quasi-linear utilities, $\{QL_t\}_{t \in \mathcal{T}=\{1,2,\dots,T\}}$, with $QL_t(q, w) = v_t(q) + w$, such that the family $\{v_t\}_{t \in \mathcal{T}}$ satisfies the well-known increasing differences condition (see, e.g., Topkis (1978) or Milgrom and Shannon (1994)). In this context, \triangleleft represents the notion of alternatives with higher quality, and it is immediate to see that every domain of menus of such objects is ordered.
- (3) **Cobb-Douglas Utility.** Consider a collection of Cobb-Douglas utilities on two-dimensional bundles, i.e., $\{CD_t\}_{t \in \mathcal{T}=\{1,2,\dots,T\}}$, with $CD_t(x_1, x_2) = x_1^{\alpha_t} x_2^{1-\alpha_t}$ such that $0 \leq \alpha_1 < \alpha_2 < \dots < \alpha_T \leq 1$. The induced relation \triangleleft trivially corresponds to the idea that: $(x_1, x_2) \triangleleft (y_1, y_2)$ if and only if $x_1 > y_1$ and $x_2 < y_2$. It is again trivial to see that, in this setting, every domain of menus of bundles is ordered.⁷

⁵For example, this is the case of multiple price lists or convex budget sets, that includes the deductible-premium insurance setting of Barseghyan, Molinari and Thirskettle (2021). Another example is the betting context of Chiappori, Salanié, Salanié and Gandhi (2019), where lotteries in a menu are defined by a possibly different prize with a possibly different probability, and a constant payoff otherwise.

⁶Alternatively, q may describe the level of leisure time in a labor application, and w the income after salary.

⁷Another standard use of this setting is for the study of other-regarding preferences where the pairs describe one's own income and the income of another individual and the order captures selfishness.

4. A CHARACTERIZATION OF \mathcal{T} -RUMs

Suppose that the analyst has access to a stochastic choice function p over an ordered domain of menus \mathcal{J} . Formally, p is 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 $\sum_{x \in A_j} p(x, j) = 1$. Consider the following property.

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

\mathcal{T} -Monotonicity captures the intuition that more support must lead to a larger choice probability. 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 alternatives in B must be lower than that of the alternatives in B' . \mathcal{T} -Monotonicity incorporates the well-known axiom of Regularity in the treatment of classical RUMs, which applies to menus related by set inclusion. In our setting, notice that if a menu is modified by the incorporation of new alternatives, the choice probability of any existing alternative cannot increase because the set of types for which it is maximal cannot expand.

An alternative way of interpreting \mathcal{T} -Monotonicity uses the largest type for which an alternative is maximal in a menu: if such largest type for alternative x in menu j is below the largest type for alternative y in menu j' , the cumulated choice probability of alternatives in j below x must be lower than that of the alternatives in j' below y .⁸ Under this version, it is immediate to see how our property also incorporates the logic of ordered RUMs described by the Centrality axiom of SCRUMs. Basically, the above reasoning implies that when considering three alternatives such that $x \triangleleft y \triangleleft z$, the elimination of alternative z from the menu does not modify the critical type of alternative x , and hence it does neither change its choice probability.

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

Theorem 1 responds to the situation where an analyst has a dataset p , and hypothesizes whether a distribution over a given ordered collection of utilities \mathcal{T} accounts for the data. Theorem 1 gives an exact answer; dataset p is rationalized by \mathcal{T} à la RUM if and only if the data satisfies \mathcal{T} -Monotonicity.

⁸It is easy to see that this alternative statement is equivalent to \mathcal{T} -Monotonicity provided that the domain includes at least one menu with no dominated options.

The following discussion is instrumental in understanding the strategy of the proof, and hence the result. First, we show that the ordered structure of menus guarantees that choices are ordered, that is, for any menu j and alternative $x \in A_j$, the set of types $\mathcal{T}(x, j)$ is always an interval and the set of types $\bigcup_{y \in A_j, y \leq x} \mathcal{T}(y, j)$ is of the form $\{1, 2, \dots, \max \mathcal{T}(x, j)\}$. That is, in every menu, choice is ordered by the set of types, with lower types selecting lower alternatives. This suggests, further, that the most relevant types are those that are the largest maximal types for the different alternatives and menus. We denote the set of such types by \mathcal{T}^I .

The proof then constructs a correspondence F on \mathcal{T}^I , that will provide the basis for the construction of the CDF over the relevant set of types that rationalizes choice. Given the ordered choices, F assigns to each type $t \in \mathcal{T}^I$ with $t = \max \mathcal{T}(x, j)$ the sum of the choice probabilities of all alternatives lower than x in menu A_j . The proof then uses \mathcal{T} -Monotonicity to show that F satisfies the conditions to be a CDF, that is, it is a single-valued increasing map, with $F(T) = 1$. Since types outside \mathcal{T}^I are inconsequential for choice, we can then construct a monotone extension, G , of F over the entire collection of types. By a monotone extension, we mean that $F(t) = G(t)$ for every $t \in \mathcal{T}^I$ and that whenever $t_1 < t_2$, $G(t_1) \leq G(t_2)$. Then, the probability distribution ψ derived from G is shown to rationalize the data.

4.1. Identification. The proof of Theorem 1 is based upon the construction of the model on the set of types \mathcal{T}^I , which are in fact the ones at which the CDF of the \mathcal{T} -RUM is fully identified. While the above discussion shows that types in this set are identified, we now discuss how, for every $t \notin \mathcal{T}^I$, the CDF at t cannot generally be fully identified. For example, consider $t_1 < t < t_2$, where t_1 and t_2 are two consecutive types in \mathcal{T}^I . If the value of the CDF at t_2 is strictly greater than its value at t_1 , the value of the CDF at t can be any value between these two. This is because, given $G(t_1) < G(t_2)$, the value $G(t) \in [G(t_1), G(t_2)]$ is irrelevant for choice, since every $t \in (t_1, t_2]$ has the same maximal alternative in every menu in the domain. Therefore, the CDF at type t can only be fully identified in the extreme case where $G(t_1) = G(t_2)$, which implies $G(t) = G(t_1) = G(t_2)$ and $\psi(t) = 0$. Obviously, data on more menus may expand the set of identifiable types. For example, in an experimental setting, the analyst may use these results to select the domain that allows the identification of the most important components of the CDF.

4.2. Complexity. We now discuss the computational complexity of testing the rationalizability of data. To do this, we use the alternative statement of \mathcal{T} -Monotonicity discussed above, that formally reads as: $\max \mathcal{T}(x, j) \leq \max \mathcal{T}(x', j') \Rightarrow \sum_{z \in A_j, z \preceq x} p(z, j) \leq \sum_{z \in A_{j'}, z \preceq x'} p(z, j')$. We argue that testing for this property is a simple task. Let $k_{\mathcal{J}} = \sum_{j \in \mathcal{J}} |A_j|$, which corresponds to the total number of possible pairs of alternatives and menus, (x, A_j) with $x \in A_j$. A brute force algorithm entails checking all such pairs, requiring a total number of $k_{\mathcal{J}}(k_{\mathcal{J}} - 1)$ checks, which is already polynomial in the input $k_{\mathcal{J}}$. This can be significantly improved, moreover, by using the following recursive argument. Let all pairs of alternatives and menus be ordered in some way. Suppose that the property holds for the first n pairs. When considering pair $n + 1$, one should relate this pair to, at most, two previously-considered pairs. For the case where there are previously considered pairs with the same largest maximal type, one can select any of these to check for the equality of their (cumulative) choice probability with that of pair $n + 1$. Otherwise, one needs to select the pairs with largest maximal types closest, above and below, to the current one, and check that the (cumulative) choice probability of pair $n + 1$ lies between the choice probabilities of these two. Hence, this algorithm involves at most $2k_{\mathcal{J}}$ comparisons, and is therefore very low in complexity.

4.3. Interval domain. 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 a generalization of this domain assumption sufficient for our purposes.

In an ordered menu j , if $t_1 < t_3$ are types leading to the choice of x , $t_2 \in (t_1, t_3)$ must also result in the same choice, given that, otherwise, the alternative chosen by t_2 would be incomparable with x according to \triangleleft . That is, the set of types $\mathcal{T}(x, j)$ is an interval set, with formal proof of this fact given in Claim 1 of Theorem 1. This leads us to consider the following domain.

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

This is indeed a larger domain, as the following example illustrates. Consider $\{x, y, z\}$, 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 , since x is maximal for types 1 and 3 and y is maximal for type 2. Hence, menus $\{x, y\}$ and $\{x, y, z\}$ are not ordered and fall outside our initial

domain assumption. However, $\{x, y, z\}$ satisfies the interval structure, since type 1 chooses x , type 2 chooses y and type 3 chooses z .

The following corollary shows that the main result also applies in this domain. We omit its proof, that follows immediately from the fact that Claim 1 in the proof of Theorem 1 follows now by assumption.

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

5. TREMBLE

Given a menu j , denote by D_j the set of alternatives that are not maximal for any of the utilities, i.e., $D_j = \{x \in A_j : \mathcal{T}(x, j) = \emptyset\}$. We know that \mathcal{T} -RUMs assign zero probability to the choice of alternatives in D_j , and their choice therefore constitutes a mistake. In this section, we extend \mathcal{T} -RUMs in order to allow for the possibility of such mistakes.

The \mathcal{T} -RUM with tremble (\mathcal{T} -RUMT) is defined by means of a possibly menu-dependent tremble function $\lambda : \mathcal{J} \rightarrow [0, 1]$, such that $\lambda_j = 0$ whenever $D_j = \emptyset$, and a probability distribution ψ over the set of types. Then, for any menu j , the total mass of choices from D_j is given by λ_j , while the choice probability of $x \in A_j \setminus D_j$ is equal to $\psi(\mathcal{T}(x, j))(1 - \lambda_j)$. That is, mistakes occur with probability λ_j , and otherwise behavior is governed by the \mathcal{T} -RUM with distribution ψ . Notice that this tremble version of the model is agnostic about the size and distribution of mistakes in D_j and allows different trembles in different menus. The only assumption in the trembling mechanism is that when choices are not mistakes, they follow a \mathcal{T} -RUM.

\mathcal{T} -RUMT is a simple model whose characterization follows immediately from the analysis in Theorem 1. Given the stochastic choice function p , denote by \bar{p} the conditional stochastic choice function:⁹

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

We then have:

Proposition 1. *In a domain of ordered menus, p is a \mathcal{T} -RUMT if, and only if, \bar{p} satisfies \mathcal{T} -Monotonicity.*

⁹When $\sum_{x \in D_j} p(x, j) = 1$, define the (irrelevant) conditional probability as $\bar{p}(x, j) = \frac{|\mathcal{T}(x, j)|}{|\mathcal{T}|}$.

Remark 1. The implementation of the tremble in \mathcal{T} -RUMT is flexible. For example, one could discard the observed choices of the non-maximal alternatives, and estimate a \mathcal{T} -RUM over the remaining data. Alternatively, one could adopt a particular structure for tremble. The simplest way would be to use a λ that is menu-independent and such that choices over non-maximal alternatives are uniformly random.

The main assumption of the trembling model is that the conditional choice probabilities over the maximal alternatives follow a \mathcal{T} -RUM. There are other plausible trembling mechanisms that would also lead to this property. For example, one could consider the possibility of mistakes occurring uniformly over the entire menu j , not just over D_j , and behaving a la \mathcal{T} -RUM otherwise.

6. ESTIMATION AND STATISTICAL TESTING

We now return to our base model described in Section 3, and present several results on the estimation and statistical testing of \mathcal{T} -RUMs when the data are finite. Formally, the data form a map $z : X \times \mathcal{J} \rightarrow \mathbb{Z}_+$, 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. The observed choice frequencies in menu j are therefore $\tilde{z}_j(\cdot) = \frac{z(\cdot, j)}{Z_j}$.

6.1. Estimation. Suppose that the data are generated by a \mathcal{T} -RUM but that, due to sampling issues, choice frequencies violate \mathcal{T} -Monotonicity. In this section, we provide a class of estimators that are based on a notion of rationalizability and show that they are strongly consistent. For this, we assume that $\mathcal{T} = \mathcal{T}^I$, that is, for ease of exposition, we work directly with the set of types that are fully identified.

Consider the following generalization of the rationalizability notion embedded in \mathcal{T} -RUMs. Let Ψ denote the set of all probability distributions over the given set of types. Let $d : \Psi \times \Psi \rightarrow \mathbb{R}_+$ be a continuous function measuring the divergence between two probability distributions such that $d(\psi, \psi) = 0$ and $|\psi - \psi'| \leq |\psi - \psi''|$ implies $d(\psi, \psi') \leq d(\psi, \psi'')$.¹⁰ Now, let $f : \mathbb{R}_+^J \rightarrow \mathbb{R}_+$ be a continuous function aggregating all deviations across menus, such that $f(0, \dots, 0) = 0$ and $\gamma = (\gamma_1, \dots, \gamma_J) \leq (\gamma'_1, \dots, \gamma'_J) = \gamma'$ implies $f(\gamma) \leq f(\gamma')$. We then say that the data z is ϵ -rationalizable if there exist

¹⁰Where \leq refers to the component wise inequality, with at least one component being different. The continuity claims in this section always involve finite real simplices or finite dimensional real spaces, and hence we use the standard Euclidean topology.

distributions $(\psi, \{\psi_j\}_{j \in J})$ such that: (i) for every j , distribution ψ_j rationalizes the choice frequencies in menu j , \tilde{z}_j , and (ii) $f(d(\psi, \psi_1), \dots, d(\psi, \psi_J)) \leq \epsilon$. That is, there exists a fundamental probability distribution, ψ , over the set of types, but at the moment of choice, the menu-dependent distribution ψ_j determines choices in menu j . However, the aggregate measure of deviations between ψ and $\{\psi_j\}_{j \in J}$ is smaller than or equal to ϵ . We can now present the following straightforward result:

Proposition 2. *For every z there is a minimum ϵ such that z is ϵ -rationalizable. Moreover, z is 0-rationalizable if, and only if, the stochastic choice function defined by its choice frequencies is a \mathcal{T} -RUM.*

With ϵ large enough, any finite choice data generated by a \mathcal{T} -RUM can be ϵ -rationalized by allowing sufficiently large menu-dependent perturbations. The continuity of maps d and f guarantees 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 magnitude of ϵ that yields ϵ -rationalizability.¹¹ Moreover, the distribution $\hat{\psi}$ in Ψ which yields the minimal value ϵ , represents an intuitive estimator, which we now show to be strongly consistent.¹²

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

The proof of the strong consistency of the estimators built on the basis of ϵ -rationalizability is related to known results of extrema estimators for a multinomial model (see, e.g., van der Vaart, 2000). The proof takes care of the fact that our model involves collections of multinomials, one per menu, the parameters of which are connected by the underlying distribution of types. Also, given that we only use the monotonicity and continuity of functions d and f , the argument applies also to non-additive estimators.

¹¹See Apesteguia and Ballester (2021) for a discussion of goodness of fit measures for stochastic choice models.

¹²Several natural examples of estimators belong to this family. In standard maximum-likelihood or least squares or minimum Chi-square estimations, 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, or the normalized square of the difference. Then, in these three standard cases, f additively aggregates all of the above-mentioned distances across menus.

6.2. Statistical Testing. We now discuss a method for statistically testing the model. Given the multinomial structure of the choices in each of the menus, we can intuitively construct the following statistic, which aggregates the standard Pearson statistic across menus in \mathcal{J} :

$$C(z) = \sum_{j \in \mathcal{J}} \sum_{x \in A_j} \frac{(z(x, j) - Z_j \psi(\mathcal{T}(x, j)))^2}{Z_j \psi(\mathcal{T}(x, j))}.$$

We can then show the following result.

Theorem 3. *$C(z)$ converges to a Chi-square with $\sum_{j \in \mathcal{J}} |A_j| - J$ degrees of freedom.*

The proof of the theorem immediately follows from the independence of the multinomial distributions across menus in a \mathcal{T} -RUM.

7. FINAL REMARKS

We close by briefly commenting on some differences between \mathcal{T} -RUMs and additive RUMs (ARUMs), which are also very popular in applied work. First, note that, in a \mathcal{T} -RUM, there is a given ordered collection of utilities over which the individual is assumed to have a preference distribution. In an ARUM, the analyst assumes that the individual has a particular utility function that is subject to additive, cardinal, shocks. This implies that, in an ARUM, the individual has a distribution with strictly positive mass over every single possible utility function over the set of alternatives, where the distribution is shaped by some assumption on the shocks (e.g., logistic, normal). Furthermore, \mathcal{T} -RUM requires only an ordinal understanding of the utility functions at stake, while ARUM requires a cardinal interpretation of the utility functions, since shocks enter additively and choice probabilities are determined by cardinal utility differences. Finally, \mathcal{T} -RUMs are monotone, in the sense that shifts in the distribution over types generate intuitive shifts in the choice distributions, thus facilitating the interpretation of the relevant behavioral parameters. However, as shown in Apestegua and Ballester (2018), the typical implementation of ARUM in risk settings, combining expected utility and i.i.d. additive errors, may suffer from severe non-monotonicity.

APPENDIX A. PROOFS

Proof of Theorem 1: The necessity of the axiom is evident and is thus omitted. We now prove its 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} -Monotonicity. The

first two claims do not use \mathcal{T} -Monotonicity; they follow exclusively from the ordered structure of the domain.

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 false. 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 $z \not\leq x$. From the joint consideration of types t_2 and t_3 , it follows that $x \not\leq z$. Given that it must be that $x \neq z$, 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 $\mathcal{T}(x, j) \neq \emptyset$, $\bigcup_{y \in A_j, y \leq x} \mathcal{T}(y, j) = \{1, 2, \dots, \max \mathcal{T}(x, j)\}$.

Proof of Claim 2: Suppose, by way of contradiction, that the claim is false. Then, there exists t^* such that either: (i) $t^* \leq \max \mathcal{T}(x, j)$ and $t^* \notin \bigcup_{y \in A_j, y \leq x} \mathcal{T}(y, j)$, or (ii) $t^* > \max \mathcal{T}(x, j)$ with $t^* \in \bigcup_{y \in A_j, y \leq x} \mathcal{T}(y, j)$ 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 $z \not\leq x$ by assumption, implying that t^* and $\max \mathcal{T}(x, j)$ must be different types. Their joint consideration guarantees, furthermore, that $x \not\leq z$, thus contradicting the fact that menu j is ordered. In case (ii), we know, by assumption, that $z \leq x$. 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. \square

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

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

$$k \in F(t) \text{ whenever there is } (x, j) \text{ such that } t = \max \mathcal{T}(x, j) \text{ and } k = \sum_{y \in A_j, y \leq x} p(y, j).$$

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

Proof of Claim 3: To see this, consider two types $t, t' \in \mathcal{T}^I$ such that $t \leq t'$. By definition of \mathcal{T}^I , there exist pairs (x, j) and (x', j') such that $t = \max \mathcal{T}(x, j)$ and $t' = \max \mathcal{T}(x', j')$.

By Claim 2, and the fact that $t \leq t'$, we know that $\bigcup_{z \in A_j, z \leq x} \mathcal{T}(z, j) = \{1, 2, \dots, t\} \subseteq \{1, 2, \dots, t'\} = \bigcup_{z \in A_{j'}, z \leq x'} \mathcal{T}(z, j')$. \mathcal{T} -Monotonicity guarantees that $\sum_{z \in A_j, z \leq x} p(z, j) \leq \sum_{z \in A_{j'}, z \leq x'} p(z, j')$, with equality when $t = t'$, which proves the claim. \square

Claim 4. \mathcal{T} -Monotonicity implies the following property, which we call \mathcal{T} -Extremeness: $p(x, j) > 0$ implies $\mathcal{T}(x, j) \neq \emptyset$.¹³

Proof of Claim 4: Let $\mathcal{T}(x, j) = \emptyset$. Hence, $\bigcup_{y \in A_j} \mathcal{T}(y, j) = \mathcal{T} \subseteq \mathcal{T} = \bigcup_{y \in A_j \setminus \{x\}} \mathcal{T}(y, j)$.

Then, by \mathcal{T} -Monotonicity, we have $\sum_{y \in A_j} p(y, j) = 1 \leq \sum_{y \in A_j \setminus \{x\}} p(y, j)$. Given that p is a stochastic choice function, $\sum_{y \in A_j \setminus \{x\}} p(y, j) \leq 1$, and, given the last inequality, it must in fact be equal to 1. Consequently, $p(x, j) = 0$, proving the claim. \square

Claim 5. $T \in \mathcal{T}^I$, with $F(T) = 1$.

Proof of Claim 5: 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 $y \preceq x$. Hence, the use of \mathcal{T} -Extremeness and the definition of stochastic choice function guarantee that $1 = \sum_{y \in A_j} p(y, j) \geq \sum_{y \in A_j, y \preceq x} p(y, j) = F(t) \geq \sum_{y \in A_j: \mathcal{T}(y, j) \neq \emptyset} p(y, j) = 1$, which proves the claim. \square

Given Claims 3 and 5, we are able to construct a map $G : \mathcal{T} \rightarrow [0, 1]$ that extends F , i.e. $F(t) = G(t)$ for every $t \in \mathcal{T}^I$, and is weakly increasing, i.e. $t_1 < t_2$ implies that $G(t_1) \leq G(t_2)$. Trivially, $\psi(t) = G(t) - G(t - 1)$, with the notational assumption $G(0) = 0$, is a probability distribution over \mathcal{T} . We then consider the \mathcal{T} -RUM defined by ψ .

Claim 6. For every menu $j \in \mathcal{J}$, and every alternative $x \in A_j$, $p(x, j) = \psi(\mathcal{T}(x, j))$.

Proof of Claim 6: If $\mathcal{T}(x, j) = \emptyset$, we know that \mathcal{T} -Monotonicity implies \mathcal{T} -Extremeness, which, in turn, 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 $\mathcal{T}(x, j) = \{\max \mathcal{T}(z, j) + 1, \max \mathcal{T}(z, j) + 2, \dots, \max \mathcal{T}(x, j)\}$ 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. \square

¹³Gul and Pesendorfer (2006) use a similar property in their study of random expected utility models.

Having constructed a \mathcal{T} -RUM that rationalizes all choice probabilities, we have proved the sufficiency of the property and hence the theorem. ■

Proof of Proposition 1: Necessity is immediate. To see sufficiency, notice that \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. ■

Proof of Proposition 2: We start by proving the first part. For a single menu $j \in \mathcal{J}$, notice that the set of distributions that rationalizes data in j is nonempty. As a result, the existence of a minimum value of ϵ such that z is ϵ -rationalizable follows directly from the compactness of $\Psi \times \Psi^J$ and the continuity of d and f . Moreover, if there exists a distribution ψ rationalizing all choices across menus, 0-rationalizability holds due to the fact that $f(d(\psi, \psi), \dots, d(\psi, \psi)) = f(0, \dots, 0) = 0$. Finally, if such a distribution does not exist, the data cannot be 0-rationalizable because of the strict monotonicity properties of both d and f . This concludes the proof. ■

Proof of Theorem 2: Given any distribution over types ψ and data frequencies \tilde{z} , consider the value $g(\psi, \tilde{z}) = \min f(d(\psi, \psi_1), \dots, d(\psi, \psi_J))$, subject to ψ_j rationalizing the choice probabilities \tilde{z}_j . Note that, using the same logic as in Proposition 2, g is well-defined. Consider a sequence of data functions $\{z^n\}_{n=1}^\infty$ with $\lim_{n \rightarrow \infty} Z_j^n = \infty$ for every $j \in \mathcal{J}$. Given the definition of the estimator and the properties of f and d , it follows that the estimator for z^n is $\hat{\psi}^n = \arg \min_{\psi \in \Psi} g(\psi, \tilde{z}^n)$.

Now suppose that the sequence of data functions is 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 $\tilde{z}^n(x, j) = 0$ always holds. For every alternative $x \in A_j$ such that $\psi^*(\mathcal{T}(x, j)) > 0$, standard arguments guarantee that the multinomial i.i.d. choices in menu j generate frequencies $\tilde{z}^n(x, j)$ that converge, almost surely, to $\psi^*(\mathcal{T}(x, j))$. Thus, the finiteness of each menu and of \mathcal{J} guarantees that \tilde{z}^n converges, almost surely, to the choice probabilities generated by ψ^* . It then follows immediately that $\hat{\psi}^n$ converges, almost surely, to ψ^* . This concludes the proof. ■

Proof of Theorem 3: As discussed in the proof of Theorem 2, the i.i.d. nature of the \mathcal{T} -RUM guarantees that we can understand choices in any given menu j as a multinomial distribution where entry x has probability $\psi(\mathcal{T}(x, j))$. It is well known

that, as the number of observations grows, the statistic $\sum_{x \in A_j} \frac{(z(x,j) - Z_j \psi(\mathcal{T}(x,j)))^2}{Z_j \psi(\mathcal{T}(x,j))}$ converges to a chi-square distribution with $|A_j| - 1$ degrees of freedom. The i.i.d. nature of the model also guarantees that, even though their parameters are linked by ψ and the domain structure is ordered, the multinomial distributions across menus are independent. Hence, it follows immediately that $C(z)$ converges to a chi-square with $\sum_{j \in \mathcal{J}} (|A_j| - 1) = \sum_{j \in \mathcal{J}} |A_j| - J$ degrees of freedom, as desired. ■

REFERENCES

- [1] Andersson, O., H.J. Holm, J.R. Tyran and E. Wengström (2020). “Robust Inference in Risk Elicitation Tasks,” *Journal of Risk and Uncertainty*, 61:195–209.
- [2] Apesteguia, J. and M.A. Ballester (2018). “Monotone Stochastic Choice Models: The Case of Risk and Time Preferences,” *Journal of Political Economy*, 126(1):74–106.
- [3] Apesteguia, J. and M.A. Ballester (2021). “Separating Predicted Randomness from Residual Behavior,” *Journal of the European Economic Association*, 19(2):1041–1076.
- [4] Apesteguia, J., M.A. Ballester and A. Gutierrez (2023). “Random Discounted Expected Utility,” mimeo.
- [5] Apesteguia, J., M.A. Ballester and J. Lu (2017). “Single-Crossing Random Utility Models,” *Econometrica*, 85(2):661–674.
- [6] Barseghyan, L., F. Molinari and M. Thirkettle (2021). “Discrete Choice under Risk with Limited Consideration,” *American Economic Review*, 111(6):1972–2006.
- [7] 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.
- [8] Cattaneo, M., X. Ma, Y. Masatlioglu and E. Suleymanov (2020). “A Random Attention Model,” *Journal of Political Economy*, 128(7):2796–2836.
- [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] Cohen, A. and L. Einav (2007). “Estimating Risk Preferences from Deductible Choice,” *American Economic Review*, 97(3):745–788.
- [11] Coller, M. and M.B. Williams (1999). “Eliciting Individual Discount Rates,” *Experimental Economics*, 2(2):107–27.
- [12] 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.
- [13] Dardanoni, V., P. Manzini, M. Mariotti and C.J. Tyson (2020). “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Econometrica*, 88(3):1269–1296.
- [14] Filiz-Ozbay, E. and Y. Masatlioglu (2023). “Progressive Random Choice,” *Journal of Political Economy*, 131(3):716–750.

- [15] Greene, W.H. and D.A. Hensher (2010). *Modeling Ordered Choices*, Cambridge University Press.
- [16] Gul, F. and W. Pesendorfer (2006). “Random Expected Utility,” *Econometrica*, 74(1):121–146.
- [17] Hammond, J.S. (1974). “Simplifying the Choice between Uncertain Prospects Where Preference is Nonlinear,” *Management Science* 20(7):1047–1072.
- [18] Jagelka, T. (2023). “Are Economists’ Preferences Psychologists’ Personality Traits? A Structural Approach,” mimeo.
- [19] Milgrom, P., and C. Shannon (1994). “Monotone Comparative Statics,” *Econometrica*, 62(1):157–180.
- [20] Petri, H. (2023). “Binary Single-Crossing Random Utility Models,” *Games and Economic Behavior*, 138:311–320.
- [21] Topkis, D.M. (1978). “Minimizing a Submodular Function on a Lattice,” *Operations Research*, 26: 305–321.
- [22] Turansick, C. (2022). “Identification in the Random Utility Model,” *Journal of Economic Theory*, 203, 105489.
- [23] Valkanova, K. (2021). “Inferring Single-Peakedness from Random Choice,” mimeo.
- [24] van der Vaart, A.W. (2000), *Asymptotic Statistics*, Cambridge University Press.
- [25] Warner, J.T. and S. Pleeter (2001). “The Personal Discount Rate: Evidence from Military Downsizing Programs,” *American Economic Review*, 91(1):33–53.