A Random Attention and Utility Model*

Nail Kashaev ⓡ Victor H. Aguiar†

May, 2021

Abstract We generalize the stochastic revealed preference methodology of McFadden and Richter (1990) for finite choice sets to settings with limited consideration. Our approach is nonparametric and requires partial choice set variation. We only impose a monotonicity condition on attention first proposed by Cattaneo et al. (2020) and a stability condition on the marginal distribution of preferences. Our framework is amenable to statistical testing. These new restrictions extend widely known parametric models of consideration.

JEL classification numbers: C50, C51, C52, C91.

Keywords: random utility, random consideration sets.

*The "ⓡ" symbol indicates that the authors' names are in certified random order, as described by Ray ⓡ Robson (2018). We gratefully acknowledge financial support from the Western Social Science Faculty Grant (FRDF R5533A02) and Social Sciences and Humanities Research Council Insight Development Grant.

†Kashaev: Department of Economics, University of Western Ontario; nkashaev@uwo.ca. Aguiar: Department of Economics, University of Western Ontario; vaguiar@uwo.ca.

1. Introduction

The stochastic revealed preference methodology of McFadden and Richter (1990) is a cornerstone of economic analysis. This research agenda aims at explaining the behavior of a population of decision makers (DMs) as if each DM maximizes her utility, which is and independent identically distributed draw from a distribution of preferences, over their choice set. This theory is usually referred to as random utility model (RUM).[1] If RUM is successful at describing behavior, then the analyst can use it to recover the distribution of heterogeneous preferences solely from observing the probability of choice of a finite set of alternatives from different menus justifying the name of revealed preference. This distribution of preference is an important input for many social sciences, and can play a key role in policy making. However, RUM may fail at describing behavior if DMs do not consider all available alternatives. This may happen, for instance, if there is a cost to understanding the decision task. In this situation, DMs may use a two-stage procedure: first simplifying choice by using a consideration set, and only then choosing the best alternative among those considered. Thus, DMs may choose dominated alternatives when facing a cost of consideration failing to be consistent with RUM.[2]

This paper proposes a generalization of the stochastic revealed preference methodology that is robust to limited consideration, allows for fully heterogeneous unrestricted preferences that can be correlated with consideration, and is amenable to statistical testing. In doing so, we provide nonparametric restrictions on limited consideration that make partial recoverability of the distribution of preferences possible in a large class of stochastic choice data sets.

A large literature, pioneered by Masatlioglu et al. (2012) and Manzini and Mariotti (2014),

[1] RUM was originally formulated by Block and Marschak (1960) and Falmagne (1978).

[2] For examples of the distortions created by limited consideration see: Ho et al. (2017) and Heiss et al. (2016) in the health insurance market. Hortaçsu et al. (2017) in the residential electricity market in Texas. Honka (2014) and Honka et al. (2017) in the US auto insurance and banking industries, respectively. De Los Santos et al. (2012) on web browsing behavior of consumers when shopping online. Barseghyan et al. (2021) on insurance purchases.

has proposed theories of consideration-mediated choice. These theories accommodate some departures from RUM caused by inattention, feasibility, categorization, and search.[3] However, most existing theories of random consideration have assumed that preferences are *homogeneous* (Cattaneo et al., 2020). This restriction means that these models are not well suited to describe behavior at the population level. Our work fills this important gap in the literature by allowing unrestricted preference heterogeneity.

The closest paper to this work is Cattaneo et al. (2020). They provide a general framework, Random Attention Model (RAM), to test different models of stochastic consideration when preferences are homogeneous. Therefore, their work is better suited for individual stochastic choice data. Cattaneo et al. (2020) impose a set-monotonicity restriction on the probability of consideration of a set of alternatives conditional on the choice set. Namely, they assume that the probability of considering a given set cannot increase if the choice set is getting larger. We study the implications of imposing this set-monotonicity constrain as well, but we allow for fully unrestricted heterogeneous preferences. We first show that RAM with heterogeneous preferences has no empirical content. Hence, in order to avoid empirical triviality, we impose a new condition on the joint distribution of preferences and consideration sets that we call *preference stability*. This condition requires that the marginal distribution of preferences does not depend on the choice set. This assumption is a generalization of RUM allowing for the presence of limited consideration.[4] Importantly, preference stability allows for statistical *dependence* between random consideration and random preferences.

Our approach differs from previous works that have used enhanced data sets to test for the presence of consideration. In particular, we only need a standard stochastic choice data set widely used in the discrete-choice literature.[5] Recently, Abaluck and Adams (2021) use

[3]See, for instance, Aguiar et al. (2016), Brady and Rehbeck (2016), Caplin et al. (2016), Aguiar (2017), Kovach and Ülkü (2020), Lleras et al. (2017), and Horan (2019).

[4]Aguiar et al. (2021) considers the same primitives but imposes full independence between random preferences and attention. In contrast to this paper, Aguiar et al. (2021) requires the presence of a default alternative and imposes parametric restrictions on the random attention.

[5]For examples of enriched data sets that identify limited consideration, see Reutskaja et al. (2011) (eye-tracking data); Honka et al. (2017) and Draganska and Klapper (2011) (additional surveys); Kawaguchi et al.

structural restrictions on the elasticity behavior of demand to identify consideration sets and preferences. We differ from that work because we do not observe attributes (e.g., prices). Barseghyan et al. (2021) obtain information about parametric distribution of preferences in a domain with attributes variation by introducing a support restriction on possible consideration sets. Our framework does not impose any parametric restrictions on the distribution of preferences and allows both shape and support restrictions on consideration probabilities. Kashaev and Lazzati (2021) develop a dynamic model of discrete choice that incorporates peer effects into random consideration sets. They identify preferences and consideration probabilities in a fixed menu settings by using variation in choices of peers. We assume menu variation and do not have access to panel data.

Aguiar ⓡ Kashaev (2021) study nonparametric identification and estimation of the distribution of consideration sets and preferences without menu variation in panel data settings. Dardanoni et al. (2020b) provide identification of the consideration probabilities given a known distribution of preferences in a fixed menu. They also consider grouped data sets where three instances of choice of the same consumers is observed to enhance identification. We assume menu variation, do not need to know the distribution of preferences, and do not use enriched stochastic choice data sets. More recently, Dardanoni et al. (2020a) provide identification arguments for both preferences and cognition heterogeneity (including consideration probabilities) in mixture data sets. In contrast to our work, their method requires observing the joint distribution of choice over different menus. Also, their results are focused on parametric heterogeneity.

The paper is organized as follows. Section 2 introduces our general framework. In Section 3, we provide its characterization and derive some implications for preference revelation. We conclude in Section 4. All proofs can be found in Appendix A.

(2016) and Conlon and Mortimer (2013) (variation in product availability); Dehmamy and Otter (2014) and Huang and Bronnenberg (2018) (variations in quantity purchased and products purchased); and Gabaix et al. (2006) (mouse-tracking data).

2. Model

Let X be a finite choice set. The collection of choice sets is denoted by a nonempty subset of the power set $\mathcal{A} \subseteq 2^X \setminus \emptyset$. We define the stochastic choice function $\rho_A \in \Delta(A)$ for $A \in \mathcal{A}$ such that $\rho_A(a)$ denotes the probability of choosing $a \in A$. The stochastic choice data set is the vector $\rho = (\rho_A)_{A \in \mathcal{A}}$.[6] We call a stochastic choice data set complete if $\mathcal{A} = 2^X \setminus \emptyset$ and incomplete otherwise.

We let $U \subseteq X \times X$ be the set of linear orders (strict preference relations) defined on X. The typical element will be denoted by $\succ \in U$. Note that U can be restricted to any subset of linear orders exhibiting some property such as single-crossing (Apesteguia et al., 2017) or expected utility. Our theory applies without changes to these restrictions. However, we present our results for the unrestricted U to maximize generality.

Within our framework, DMs may exhibit limited consideration. DMs exhibit limited consideration when they maximize their preferences in a strict subset of the observed choice set. This strict subset is called a consideration set. We model limited consideration using the notion of consideration filters.

Definition 1 (Consideration Filter)**.** Given a set $D \in 2^X \setminus \emptyset$ we say that $\Gamma_D : 2^X \setminus \emptyset \to 2^X$ is a feasible consideration filter if

$$\Gamma_D(A) = \begin{cases} D, & D \subseteq A, \\ \emptyset, & \text{otherwise} \end{cases}$$

for all $A \in 2^X \setminus \emptyset$.

Let $\Phi = \bigcup_{D \in 2^X \setminus \emptyset} \{\Gamma_D \,:\, \Gamma_D \text{ is a feasible filter}\}$ be a finite collection of all feasible filters. The typical element of it will be denoted by $\phi \in \Phi$.

[6] Since most dataset do not collect data on singleton menus, we assume for these cases that $\rho_{\{a\}}(a) = 1$ for all $a \in X$.

We consider a random attention and utility model (RAUM). A *behavioral type* of this model is determined by a pair of preferences and filter $(\succ, \phi) \in U \times \Phi$. A RAUM rule $\pi = (\pi_A)_{A \in 2^X \setminus \emptyset}$ is a collection of probability distribution over preferences and filters $\pi_A \in \Delta(U \times \Phi)$ for all choice sets $A \in 2^X \setminus \emptyset$. This rule has the property that $\pi_A(\succ, \phi) = 0$ whenever $\phi(A) = \emptyset$.

Definition 2. A stochastic choice data set ρ admits a RAUM representation π if

$$\rho_A(a) = \sum_{(\succ,\phi) \in U \times \Phi} \pi_A(\succ, \phi) \mathbb{1}\left(a \in \phi(A),\ a \succ b,\ \forall b \in \phi(A) \right),$$

for all $a \in A$ and all $A \in \mathcal{A}$.

The RAUM is so general that it does not have any empirical content. That is, feasible filters are permissive enough to explain any behavior. Hence, without further constraints, it is impossible to falsify the RAUM or to recover the (marginal) distribution of preferences of a population of DMs $\pi^*_A(\succ) = \sum_\phi \pi_A(\succ, \phi)$. We impose the following stability constraint on the RAUM representation π.

Assumption 1 (Stability)**.** There exists $\pi^* \in \Delta(U)$ such that $\pi^*_A(\succ) = \pi^*(\succ)$ for any $A \in 2^X \setminus \emptyset$ and $\succ \in U$.

Note that stability is equivalent to requiring that $\pi^*_A(\succ) = \pi^*_B(\succ)$ for any $A, B \in \mathcal{A}$ and $\succ \in U$, thus, justifying its name.

One interpretation of stability is that it restricts limited consideration such that the marginal distribution of preferences of the general RAUM is equivalent to the *true* distribution of heterogeneous preferences in the population. The true distribution of preferences is the distribution on $\Delta(U)$ that controls behavior in the counterfactual situation of absence of limited consideration. Of course, our stability assumption does not require the knowledge of such distribution. Moreover, our notion of stability of the distribution of preferences is a natural generalization of the assumption of preference stability in the stochastic rationality model of McFadden and Richter (1990).

Importantly, stability allows for stochastic dependence between consideration filters and random preferences (see Example 7). We only require that the (marginal) distribution of preferences remains the same across exogenously given menus of alternatives.

Even under stability, limited consideration has to be further restricted to have empirical bite as we will show in Proposition 1. Here, we follow Cattaneo et al. (2020) and impose the following restriction.

Assumption 2 (Set-monotonicity)**.** For any $\succ$, ϕ, A, and B such that $A \subseteq B$ and $\phi(A) \neq \emptyset$, it must be that $\pi_A(\phi| \succ) \geq \pi_B(\phi| \succ)$.

Set-monotonicity means that the conditional probability of a given filter, ϕ, conditional on a preferences type, $\succ$, cannot increase as the menu expands. That is, DMs will pay more attention to a set when the menu of alternatives is smaller. Intuitively, larger menus have a higher opportunity cost of consideration. Cattaneo et al. (2020) show that many models of random consideration satisfy set-monotonicity.

The next proposition qualifies the importance of stability and set-monotonicity working together. Neither of these restrictions alone are enough for empirical relevance of the model. However, when they are combined together, the model becomes falsifiable even with limited menu variation. We see the combination of these two restrictions as a baseline of empirical content that makes our study empirically meaningful.

Proposition 1. The following statements are true:

(i) Any ρ admits a stable RAUM representation.

(ii) Any ρ admits a set-monotone RAUM representation.

(iii) There exists an incomplete ρ that does not admit a set-monotone stable RAUM representation.

Here we provide a sketch of the proof of (iii). We construct an incomplete data set (i.e., $\mathcal{A} \neq 2^X \setminus \emptyset$) that does not admit a set-monotone stable RAUM. Let $X = \{a, b, c, d\}$ and

$$\mathcal{A} = \{\{a, b\}, \{a, c\}, \{b, d\}\{a, b, d\}, \{a, c, d\}, \{b, c, d\}\}.$$

Suppose the observed ρ is as follows

	$\{a, b\}$	$\{a, c\}$	$\{b, d\}$	$\{a, b, d\}$	$\{a, c, d\}$	$\{b, c, d\}$
$\rho_A(a)$	1	1	0	0	0	0
$\rho_A(b)$	0	0	1	1	0	α_b
$\rho_A(c)$	0	0	0	0	1	α_c
$\rho_A(d)$	0	0	0	0	0	α_d

where $\alpha_d > 0$. Consider the pair $\{a, b\}$ and $\{a, b, d\}$. From observing $\rho_{\{a,b\}}(b) = 0$ and $\rho_{\{a,b,d\}}(b) = 1$, we can conclude that $b \succ d$ with probability 1 or b and d are never considered together. Similarly, from observing $\rho_{\{a,c\}}(c) = 0$ and $\rho_{\{a,c,d\}}(c) = 1$, we can make analogous conclusion about c and d. As a result, the fact that $\alpha_d > 0$ then implies that the probability of considering the singleton consideration set $\{d\}$ must be nonzero (otherwise d is either never considered or dominated by b or c). But the latter is impossible because in menu $\{b, d\}$ option d is never chosen. The formal details of the sketch above can be found in Appendix A.1.

Now we introduce several examples showing that RAUM generalizes a large variety of models of interest in the literature of stochastic rationality and random consideration.

Example 1 (Random Utility, RUM, Block and Marschak, 1960)**.** Let $\pi_A(\phi| \succ) = \mathbb{1}\,(\,\phi(A) = A\,)$ for all $A \in 2^X \setminus \emptyset$. That is, the whole menu is always considered. Then the marginal preference rule will induce a RUM such that for each $A \in \mathcal{A}$

$$\rho_A(a) = \sum_{\succ} \pi^*(\succ) \mathbb{1}\,(\,a \succ b,\; \forall b \in A\,)\,.$$

Example 2 (Random Attention Model, RAM, Cattaneo et al., 2020)**.** Let the marginal distribution of preferences be degenerate. That is, $\pi(\succ^*) = 1$ for some preferences $\succ^*$. Then the rule $\pi_A(\phi|\succ^*)$ will induce a RAM such that for each $A \in \mathcal{A}$

$$\rho_A(a) = \sum_{\phi} \pi_A(\phi|\succ^*)\mathbb{1}\left(a \in \phi(A),\ a \succ^* b,\ \forall b \in \phi(A)\right).$$

Given $\succ$, an attention-index $\eta_\succ : 2^X \to \mathbb{R}_+$ such that $\eta_\succ(\emptyset) = 0$ and $\sum_{C \subseteq X} \eta_\succ(C) = 1$ is a probability measure over subsets of X. The following models of consideration are examples of rules that are governed by attention-indexes.

Example 3 (Preference Dependent Logit Attention, LA, Brady and Rehbeck, 2016)**.** For every $\succ$ let $\eta_\succ$ be an attention-index. Then

$$\pi_A(\phi|\succ) = \frac{\eta_\succ(\phi(A))}{\sum_{\phi'} \eta_\succ(\phi'(A))},$$

will form a preference dependent logit attention model.

Example 4 (Item Specific Attention, MM, Manzini and Mariotti, 2014, Horan, 2019)**.** MM is an LA model where $\eta_\succ$ satisfies

$$\eta_\succ(A) = \frac{\prod_{a \in X \setminus A}(1 - \gamma_\succ(a)) \prod_{b \in A} \gamma_\succ(b)}{1 - \prod_{c \in X}(1 - \gamma_\succ(c))}$$

for some $\gamma_\succ : X \to (0, 1)$, for all $A \in 2^X$ and $\succ \in U$ with the convention that $\prod_{b \in \emptyset} \gamma_\succ(b) = 0$.

Example 5 (Elimination by Aspects, EBA, Tversky, 1972)**.** For every $\succ$ take any attention-index $\eta_\succ$. Then

$$\pi_A(\phi|\succ) = \sum_{C \subseteq X\,:\, C \cap A = \phi(A)} \frac{\eta_\succ(C)}{\sum_{B \subseteq X\,:\, B \cap A \neq \emptyset} \eta_\succ(B)},$$

for all ϕ such that $\phi(A) \neq \emptyset$ will form an elimination by aspects model.

Sometimes we have an alternative that is always present in $\mathcal{A}$. That is, there exists o such that $o \in A$ for all $A \in \mathcal{A}$. We call this the default alternative.

Example 6 (MM with default, Manzini and Mariotti, 2014)**.** Item specific attention model with default is an MM model with a default such that $\gamma_{\succ}(o) = 1$ for all $\succ$ and o is the worst item with probability 1.[7]

3. Characterization of Set-monotone and Stable RAUM and Preference Revelation

In this section we characterize the set-monotone and stable RAUM in a form amenable to (statistical) testing. In particular, we show that to conclude whether a given data set admits a set-monotone and stable RAUM representation, it suffices to check whether a particular linear program has a solution. This problem is similar to the one in McFadden and Richter (1990) that characterizes RUM.

Let d_ρ, d_m, and d_r denote the number of entries in ρ, the cardinality of $2^X \setminus \emptyset$, and the total number of linear restrictions imposed by feasibility, stability, and set-monotonicity on π, respectively. Also, define $g = (\rho', 1'_{d_m}, 0'_{d_r})' \in \mathbb{R}^{d_g}$, where 1_{d_m} is the vector of ones of length d_m and 0_{d_r} is the vector of zeros of length d_r.

Theorem 1. *Given a stochastic choice data set ρ the following are equivalent:*

(i) ρ admits a set-monotone stable RAUM.

(ii) There exists $v \in \mathbb{R}^d_+$, $d < \infty$, such that

$$g = Gv$$

where G is a known matrix that consists of -1, 0, 1.

[7] The EBA is also called Random Categorization Rule (RCG) (Aguiar, 2017) when there is a default alternative that always attracts attention. Suleymanov (2018) shows that MM with a default is the intersection of LA and EBA/RCG with a default.

Theorem 1 provides a linear characterization of a set-monotone stable RAUM. It is important to note that without stability the problem is quadratic since set-monotonicity is imposed on the conditional distribution over filters $\pi_A(\phi|\succ)$. Linearity of our problem is amenable to statistical testing using tools in Deb et al. (2018) as we discuss below.

To better understand how matrix G is constructed, we sketch the proof of Theorem 1. The formal proof and the detailed description of G and g can be found in Appendix A.2. According to the definition of the RAUM representation

$$\rho_A(a) = \sum_{(\succ,\phi)\in U\times\Phi} \pi_A(\succ,\phi)\mathbb{1}\left(\, a\in\phi(A),\ a\succ b,\ \forall b\in\phi(A)\,\right).$$

Thus, if we stack all (including unobserved menus) $\pi_A(\succ,\phi)$s into a vector $\tilde{v}\in\mathbb{R}_+^{\left|2^X\setminus\emptyset\right|\cdot|U|\cdot|\Phi|}$, then we can represent every $\rho_A(a)$ is an inner product of $\tilde{v}$ and some vector that consists of zeros and ones. If we create a matrix B such that every row of B is one of these 0-1-vectors transposed, then we will have a system of linear equations $\rho = B\tilde{v}$. Note that the columns of B that correspond to menus that are not in $\mathcal{A}$ are zero columns. This is the only set of linear equations that connects the observed (estimated) ρ and the model parameters $\tilde{v}$. Without any additional restrictions testing whether there exists component-wise nonnegative $\tilde{v}$ such that $\rho = B\tilde{v}$ can be done on the basis of Kitamura and Stoye (2018) since only ρ needs to be estimated. However, as Proposition 1 shows, we need to impose monotonicity and stability to have empirical bite. These additional restrictions would require us to use a testing procedure presented in Deb et al. (2018).

The rest of the constraints are linear equality constraints on parameters and do not use data. First, we need to impose the condition that $\sum_{\succ,\phi}\pi_A(\succ,\phi) = 1$ for all $A\in 2^X\setminus\emptyset$. This constraints are linear and can be written as $1_{d_m} - O\tilde{v} = 0$ for matrix O that consists of -1,0, and 1.

Next we need to impose the feasibility condition that requires $\pi_A(\succ,\phi)$ to be zero whenever $\phi(A)=\emptyset$. These constraints are linear and can be written as $F\tilde{v}=0$, where F is with only a single 1 in every row and column.

To impose stability, note that it can be stated as

$$\sum_{\phi} \pi_A(\succ, \phi) - \pi_B(\succ, \phi) = 0$$

for every $A, B \in \mathcal{A}$. Hence, in the matrix form it can be rewritten as $S\tilde{v} = 0$, where S is the matrix that consists of -1, 0, and 1.

To impose set-monotonicity, note that under stability for $A \subseteq B$

$$\pi_A(\phi | \succ) \geq \pi_B(\phi | \succ)$$

is equivalent to

$$\pi_A(\phi | \succ)\pi^*(\succ) = \pi_A(\succ, \phi) \geq \pi_B(\succ, \phi) = \pi_B(\phi | \succ)\pi^*(\succ).$$

Hence, we can construct a vector of nonnegative slack variables $\bar{v}$ such that the set-monotonicity constraints can be written as $Mv = 0$, where M is the matrix that consists of -1, 0, 1, and $v = (\tilde{v}', \bar{v}')'$. Finally, define

$$G = \begin{bmatrix} B & 0 \\ O & 0 \\ F & 0 \\ S & 0 \\ M & -I \end{bmatrix},$$

where I is the identity matrix of proper size. As a result, we can summarize a set-monotone stable RAUM as $g = Gv$, where the first $\sum_{A \in \mathcal{A}} |A|$ elements of g are those of ρ, the next block of g correspond to 1_{d_m}, and the rest are zeros. From statistical testing prospective, the above setting is covered by the testing procedure proposed in Deb et al. (2018).

The main result in this section generalizes Cattaneo et al. (2020) to heterogeneous preferences.

We highlight that our approach is different from Cattaneo et al. (2020) as we use feasible filters instead of triangular filters. Even when the latter help reducing dimensionality, triangular filters are tailor-made for set-monotonicity and homogeneous preferences. In addition, recent methodological advances in random column generation in Smeulders et al. (2021) allow us to handle the potentially large dimensionality of our problem.

Another key difference from Cattaneo et al. (2020) is that we do not require the data set to be complete. We show that verifying the conditions on Theorem 1 is necessary and sufficient to guarantee that there is a set-monotone stable RAUM representation of the data set.

We conclude this section by providing a simple example of a set-monotone stable RAUM representation where preferences and consideration are not independent. This illustrates that the interaction of both conditions does not imply independence of consideration and preferences.

Example 7. Let $X = \{a, b\}$ and $\mathcal{A} = \{\{a\}, \{b\}, \{a, b\}\}$. Let $a \succ_1 b$ and $b \succ_2 a$, and assume that only two filters below realize with nonzero probability:

$$\phi_1(A) = \begin{cases} \{a\}, & \text{if } A \in \{\{a\}, \{a, b\}\}, \\ \emptyset, & \text{otherwise,} \end{cases}$$

and

$$\phi_2(A) = \begin{cases} \{b\}, & \text{if } A \in \{\{b\}, \{a, b\}\}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Consider the following distributions over the above two preference orders and two filters:

$$\pi_{\{a,b\}}(\succ, \phi) = \begin{cases} 1/3, & \text{if } (\succ, \phi) = (\succ_1, \phi_1), \\ 1/6, & \text{if } (\succ, \phi) = (\succ_1, \phi_2), \\ 1/6, & \text{if } (\succ, \phi) = (\succ_2, \phi_1), \\ 1/3, & \text{if } (\succ, \phi) = (\succ_2, \phi_2), \end{cases}$$

$$\pi_{\{a\}}(\succ, \phi) = \begin{cases} 1/2, & \text{if } (\succ, \phi) = (\succ_1, \phi_1), \\ 0, & \text{if } (\succ, \phi) = (\succ_1, \phi_2), \\ 1/2, & \text{if } (\succ, \phi) = (\succ_2, \phi_1), \\ 0, & \text{if } (\succ, \phi) = (\succ_2, \phi_2), \end{cases}$$

and

$$\pi_{\{b\}}(\succ, \phi) = \begin{cases} 0, & \text{if } (\succ, \phi) = (\succ_1, \phi_1), \\ 1/2, & \text{if } (\succ, \phi) = (\succ_1, \phi_2), \\ 0, & \text{if } (\succ, \phi) = (\succ_2, \phi_1), \\ 1/2, & \text{if } (\succ, \phi) = (\succ_2, \phi_2). \end{cases}$$

Note that $\pi_A(\succ_1) = \pi_A(\succ_2) = 1/2$ for all A. However,

$$\frac{1}{3} = \pi_{\{a,b\}}(\succ_1, \phi_1) \neq \pi_{\{a,b\}}(\succ_1) \sum_{i=1}^{2} \pi_{\{a,b\}}(\succ_i, \phi_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

That is, preferences and filters are not independent. Moreover, set-monotonicity is also satisfied. For instance,

$$\pi_{\{a\}}(\phi_1 | \succ_1) = \frac{1/2}{1/2} = 1 \geq \frac{2}{3} = \frac{1/3}{1/2} = \pi_{\{a,b\}}(\phi_1 | \succ_1).$$

Partial Identification of Preferences

Although RAUM is falsifiable, given that preferences are not homogeneous, it is important to learn whether RAUM reveals anything about preferences. In this section, we show that RAUM reveals information about the distribution of preferences in population in some datasets.

We say that ρ is regular if $\rho_A(a) \geq \rho_B(a)$ for all $A \subseteq B$ and $a \in A$. Otherwise, we call ρ irregular.

To formalize the notion of revelation of preferences, let R_ρ be a set of all set-monotone stable RAUM representations of ρ. That is,

$$\mathrm{R}_\rho = \{\pi \,:\, \rho \text{ admits set-monotone stable RAUM } \pi\}\,.$$

Next define the identified set for preference distributions implied by R_ρ as

$$\Pi(\mathrm{R}_\rho) = \left\{\pi^* \,:\, \pi^*(\succ) = \sum_{\phi} \pi_A(\succ, \phi) \text{ for all } A, \succ \text{ and some } \pi \in \mathrm{R}_\rho\right\}.$$

Proposition 2. $\Pi(\mathrm{R}_\rho)$ is a strict subset of $\Delta(U)$ for any irregular ρ.

Proposition 2 states that irregular data is always informative about preferences. Since set-monotone stable RAUM is a generalization of RAM, the conclusion of Proposition 2 is a generalization of the results in Cattaneo et al. (2020) for heterogeneous preferences.[8]

[8] Note that $\Pi(\mathrm{R}_\rho)$ and $\Delta(U)$ are closed sets, hence, the difference between $\Delta(U)$ and $\Pi(\mathrm{R}_\rho)$ has a positive Lebesgue measure.

4. Conclusions

We have extended the classical stochastic revealed preference methodology in McFadden and Richter (1990) for finite sets to allow for limited consideration. Our model allows for heterogeneous preferences that are correlated with consideration sets. Fully unrestricted correlation between preferences and consideration render the set-monotonicity restriction on consideration in Cattaneo et al. (2020) void of empirical content. We introduce a new condition that restricts the relation between preferences and consideration minimally, called preference stability that reestablished the empirical bite of this model. The proposed model and conditions are amenable to statistical testing using procedures proposed in Deb et al. (2018) and Kalouptsidi et al. (2020).

References

Abaluck, Jason and Abi Adams (2021) "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, Accepted. 1

Aguiar, Victor H (2017) "Random categorization and bounded rationality," *Economics Letters*, 159, 46–52. 3, 7

Aguiar, Victor H., Maria Jose Boccardi, and Mark Dean (2016) "Satisficing and stochastic choice," *Journal of Economic Theory*, 166, 445 – 482, https://doi.org/10.1016/j.jet.2016.08.008. 3

Aguiar, Victor H, Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim (2021) "Random Utility and Limited Consideration." 4

Aguiar, Victor H ⓡ Nail Kashaev (2021) "Identification and Estimation of Discrete Choice Models with Unobserved Choice Sets." 1

Apesteguia, Jose, Miguel A Ballester, and Jay Lu (2017) "Single-Crossing Random Utility Models," *Econometrica*, 85 (2), 661–674. 2

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum (2021) "Heterogeneous choice sets and preferences," *Econometrica*, forthcoming. 2, 1

Block, Henry David and Jacob Marschak (1960) "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, 2, 97–132. 1, 1

Brady, Richard L and John Rehbeck (2016) "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84 (3), 1203–1223. 3, 3

Caplin, Andrew, Mark Dean, and John Leahy (2016) "Rational inattention, optimal consideration sets and stochastic choice,"Technical report, Working paper. 3

Cattaneo, Matias D, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov (2020) "A random attention model," *Journal of Political Economy*, 128 (7), 2796–2836. (document), 1, 2, 2, 2, 3, 3, 4

Conlon, Christopher T and Julie Holland Mortimer (2013) "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 5 (4), 1–30. 5

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, Henrik Petri, and Christopher J Tyson (2020a) "Mixture Choice Data: Revealing Preferences and Cognition." 1

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J Tyson (2020b) "Inferring cognitive heterogeneity from aggregate choices," *Econometrica*, 88 (3), 1269–1296. 1

De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012) "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102 (6), 2955–80, 10.1257/aer.102.6.2955. 2

Deb, Rahul, Yuichi Kitamura, John K-H Quah, and Jörg Stoye (2018) "Revealed price preference: theory and empirical analysis," *arXiv preprint arXiv:1801.02702*. 3, 4

Dehmamy, Keyvan and Thomas Otter (2014) "Utility and attention-a structural model of consideration," *Available at SSRN*. 5

Draganska, Michaela and Daniel Klapper (2011) "Choice set heterogeneity and the role of advertising: An analysis with micro and macro data," *Journal of Marketing Research*, 48 (4), 653–669. 5

Falmagne, J. C. (1978) "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18 (1), 52–72, 10.1016/0022-2496(78)90048-2. 1

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006) "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, 96 (4), 1043–1068. 5

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou (2016) "Inattention and switching costs as sources of inertia in medicare part d,"Technical report, National Bureau of Economic Research. 2

Ho, Kate, Joseph Hogan, and Fiona Scott Morton (2017) "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 48 (4), 877–905. 2

Honka, Elisabeth (2014) "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45 (4), 847–884, 10.1111/1756-2171.12073. 2

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (2017) “Advertising, consumer awareness, and choice: Evidence from the us banking industry,” *The RAND Journal of Economics*, 48 (3), 611–646. 2, 5

Horan, Sean (2019) “Random consideration and choice: A case study of "default" options,” *Mathematical Social Sciences*, 102, 73–84. 3, 4

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L Puller (2017) “Power to choose? An analysis of consumer inertia in the residential electricity market,” *American Economic Journal: Economic Policy*, 9 (4), 192–226. 2

Huang, Yufeng and Bart J Bronnenberg (2018) “Pennies for your thoughts: Costly product consideration and purchase quantity thresholds,” *Marketing Science*. 5

Kalouptsidi, Myrto, Yuichi Kitamura, Lucas Lima, and Eduardo A Souza-Rodrigues (2020) “Partial identification and inference for dynamic models and counterfactuals,”Technical report, National Bureau of Economic Research. 4

Kashaev, Nail and Natalia Lazzati (2021) “Peer Effects in Random Consideration Sets,” *arXiv preprint arXiv:1904.06742*. 1

Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe (2016) “Identifying Consumer Attention: A Product-Availability Approach.” 5

Kitamura, Yuichi and Jörg Stoye (2018) “Nonparametric analysis of random utility models,” *Econometrica*, 86 (6), 1883–1909. 3

Kovach, Matthew and Levent Ülkü (2020) “Satisficing with a variable threshold,” *Journal of Mathematical Economics*, 87, 67–76. 3

Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay (2017) “When more is less: Limited consideration,” *Journal of Economic Theory*, 170, 70–85. 3

Manzini, Paola and Marco Mariotti (2014) "Stochastic choice and consideration sets," *Econometrica*, 82 (3), 1153–1176. 1, 4, 6

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012) "Revealed Attention," *American Economic Review*, 102 (5), 2183–2205, 10.1257/aer.102.5.2183. 1

McFadden, Daniel and Marcel K Richter (1990) "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186. (document), 1, 2, 3, 4

Ray, Debraj ⓡ Arthur Robson (2018) "Certified random: A new order for coauthorship," *American Economic Review*, 108 (2), 489–520. *

Reutskaja, Elena, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel (2011) "Search dynamics in consumer choice under time pressure: An eye-tracking study," *American Economic Review*, 101 (2), 900–926. 5

Smeulders, Bart, Laurens Cherchye, and Bram De Rock (2021) "Nonparametric analysis of random utility models: computational tools for statistical testing," *Econometrica*, 89 (1), 437–455. 3

Suleymanov, Elchin (2018) "Stochastic attention and search." 7

Tversky, Amos (1972) "Elimination by aspects: A theory of choice.," *Psychological review*, 79 (4), 281. 5

A. Proofs

A.1. Proof of Proposition 1

Since any ρ can be completed, it is sufficient to establish validity of statements (i) and (ii) for complete stochastic data sets (i.e., $\mathcal{A} = 2^X \setminus \emptyset$).

Proof of (i). Fix any complete ρ and let $\pi_A(\phi| \succ) = \rho_A(a)\mathbb{1}\left(\phi(A) = \{a\} \right)$ for all $a, A, \succ$. Then ρ admits a stable RAUM representation $\pi_A(\phi| \succ)\pi^*(\succ)$, where π^* is any element in $\Delta(U)$.

Proof of (ii). Fix any complete ρ. For any $\succ, A$ let $a_{\succ,A}$ be the best element in A according to $\succ$ and $\kappa_{\succ,A} = \sum_{\succ' \in U} \mathbb{1}\left(a_{\succ,A} = a_{\succ',A} \right)$ be the number of preference orders for which $a_{\succ,A}$ is also the best. Take $\pi_A(\succ) = \rho_A(a_{\succ,A})/\kappa_{\succ,A}$. Then ρ admits a monotone RAUM representation $\pi_A(\phi| \succ)\pi_A(\succ)$, where $\pi_A(\phi| \succ) = \mathbb{1}\left(\phi(A) = A \right)$ for all $A, \succ$.

Proof of (iii). To prove (iii) we will construct an incomplete data set (i.e., $\mathcal{A} \neq 2^X \setminus \emptyset$) that does not admit a monotone stable RAUM. Let $X = \{a, b, c, d\}$ and

$$\mathcal{A} = \{\{a, b\}, \{a, c\}, \{b, d\}\{a, b, d\}, \{a, c, d\}, \{b, c, d\}\}.$$

Suppose the observed ρ is as follows

	$\{a,b\}$	$\{a,c\}$	$\{b,d\}$	$\{a,b,d\}$	$\{a,c,d\}$	$\{b,c,d\}$
$\rho_A(a)$	1	1	0	0	0	0
$\rho_A(b)$	0	0	1	1	0	α_b
$\rho_A(c)$	0	0	0	0	1	α_c
$\rho_A(d)$	0	0	0	0	0	α_d

where $\alpha_d > 0$. Consider the pair $\{a, b\}$ and $\{a, b, d\}$. Note that

$$
\begin{aligned}
0 = \rho_{\{a,b\}}(b) &= \sum_{\succ} \pi_{\{a,b\}}(\succ, \{b\}) + \pi_{\{a,b\}}(\succ, \{a,b\}) \mathbb{1}\,(\, b \succ a\,)\,, \\
1 = \rho_{\{a,b,d\}}(b) &= \sum_{\succ} \pi_{\{a,b,d\}}(\succ, \{b\}) + \pi_{\{a,b,d\}}(\succ, \{a,b\}) \mathbb{1}\,(\, b \succ a\,) \\
&\qquad + \pi_{\{a,b,d\}}(\succ, \{b,d\}) \mathbb{1}\,(\, b \succ d\,) + \pi_{\{a,b,d\}}(\succ, \{a,b,d\}) \mathbb{1}\,(\, b \succ a, d\,)\,.
\end{aligned}
$$

Subtracting the first equation from the second one, we get that

$$
\begin{aligned}
1 = &\sum_{\succ} [\pi_{\{a,b,d\}}(\succ, \{b\}) - \pi_{\{a,b\}}(\succ, \{b\})] + [\pi_{\{a,b,d\}}(\succ, \{a,b\}) - \pi_{\{a,b\}}(\succ, \{a,b\})] \mathbb{1}\,(\, b \succ a\,) \\
&+ \pi_{\{a,b,d\}}(\succ, \{b,d\}) \mathbb{1}\,(\, b \succ d\,) + \pi_{\{a,b,d\}}(\succ, \{a,b,d\}) \mathbb{1}\,(\, b \succ a, d\,)\,.
\end{aligned}
$$

Set-monotonicity of π_A and stability of preferences then imply that

$$
1 \leq \sum_{\succ} \pi_{\{a,b,d\}}(\succ, \{b,d\}) \mathbb{1}\,(\, b \succ d\,) + \pi_{\{a,b,d\}}(\succ, \{a,b,d\}) \mathbb{1}\,(\, b \succ a, d\,)\,.
$$

Since $\sum_{\succ} \sum_{D \subseteq \{a,b,d\}} \pi_{\{a,b,d\}}(\succ, D) = 1$, we can conclude that $\pi(\succ)$ is such that $b \succ d$ with probability 1. If we apply the above arguments to $\rho_{\{a,c\}}(c)$ and $\rho_{\{a,c,d\}}(c)$, we can deduce that $c \succ d$ with probability 1. Thus, with probability 1, d is never picked if it is considered together with b or c. Hence, in menu $\{b, c, d\}$ it can be picked with positive probability if and only if set $\{d\}$ is considered with positive probability. The later is not possible since $\pi_{\{b,c,d\}}(\succ, \{d\}) \leq \pi_{\{b,d\}}(\succ, \{d\}) \leq \rho_{\{b,d\}}(d)$ and d is never picked in menu $\{b, d\}$.

A.2. Proof of Theorem 1

Assume that π is a set-monotone stable representation of possibly incomplete ρ. Let $d_m = \left|2^X \setminus \emptyset \times U \times \Phi\right|$ and $\mathcal{A}_a = \{(a, A) \in X \times \mathcal{A} \,:\, a \in A\}$. Fix one-to-one mapping $i_1 : \mathcal{A}_a \to \{1, 2, \ldots, |\mathcal{A}_a|\}$ that maps a pair (a, A) to a corresponding element of vector ρ. Also fix any one-to-one $i_2 : 2^X \setminus \emptyset \times U \times \Phi \to \{1, 2, \ldots, d_m\}$. Let B be a matrix of size $|\mathcal{A}_a| \times d_m$ such that the (k, l)-element of it, $B_{k,l}$, is defined as follows

$$B_{k,l} = \begin{cases} \mathbb{1}\left(a \succ b, \forall b \in \phi(A) \right), & \text{if } k = i_1((a, A)), l = i_2(A, \succ, \phi) \text{ for some } \succ, \phi, (a, A) \in \mathcal{A}_a \\ 0, & \text{otherwise} \end{cases}.$$

Hence, in matrix notation, if ρ admits a RAUM representation, then

$$\rho = B\pi,$$

where $\pi = (\pi_A(\succ, \phi))_{i_2(A,\succ,\phi)}$.

The rest of the restrictions will be imposed on *all* menus (including the ones that are not present in $\mathcal{A}$). These restrictions do not use any data. First, we want to capture the fact that $\pi_A(\cdot, \cdot)$ is a probability distribution and needs to sum up to 1. For any $A \in 2^X \setminus \emptyset$, let an $i_2(A, \succ, \phi)$ element of a row of matrix O to be 1 for all $\succ$ and ϕ and to be zero otherwise. Hence, the constraint can be written as

$$O\pi = 1_{d_m},$$

where O is the matrix of size $d_m \times d_1$.

The next set of restrictions captures feasibility: $\pi_A(\succ, \phi) = 0$ whenever $\phi(A) = \emptyset$. Let $d_2 = \sum_{A,\succ,\phi} \mathbb{1}\left(\phi(A) = \emptyset \right)$. The the feasibility constraint can be written as

$$F\pi = 0,$$

where F is a matrix of 0/1 that picks $i_2(A, \succ, \phi)$ elements of π that should be set to zero because of feasibility.

Next we want to rewrite the definition of stability in the matrix form. Note that stability can be written as $\sum_\phi \pi_A(\succ, \phi) = \sum_\phi \pi_B(\succ, \phi)$ for all A, B. Fix any A, B and $\succ$. Let $\iota^{A,B,\succ}$ be a vector of length d such that

$$\iota_k^{A,B,\succ} = \mathbb{1}\left(\exists \phi \,:\, k = i_2(A, \succ, \phi)\right) - \mathbb{1}\left(\exists \phi \,:\, k = i_2(B, \succ, \phi)\right).$$

Take a collection of vectors $\left\{\iota_k^{A,B,\succ}\right\}_{A,B,\succ}$ and remove all linearly dependent or zero vectors. Let every element of what is left to be a row of a matrix S. Then, stability is equivalent to

$$S\pi = 0.$$

Finally, we want to build a matrix representation of set-monotonicity. Note that, under stability, $\pi_A(\phi|\succ) \geq \pi_B(\phi|\succ)$ is equivalent to $\pi_A(\succ, \phi) \geq \pi_B(\succ, \phi)$. Hence, similarly to stability, fix any $A, B, \succ, \phi$ such that $A \subseteq B$, $A \neq B$, and let $\iota^{A,B,\succ,\phi}$ be a vector of length d such that

$$\iota_k^{A,B,\succ,\phi} = \mathbb{1}\left(k = i_2(A, \succ, \phi)\right) - \mathbb{1}\left(k = i_2(B, \succ, \phi)\right).$$

Similarly to matrix S we can use vectors $\left\{\iota_k^{A,B,\succ,\phi}\right\}$ to build matrix M such that set-monotonicity is equivalent to

$$M\pi = \bar{v},$$

where $\bar{v}$ is a component-wise nonnegative vector. Define G as

$$G = \begin{bmatrix} B & 0 \\ O & 0 \\ F & 0 \\ S & 0 \\ M & -I \end{bmatrix}.$$

As a result, if ρ admits a set-monotone stable RAUM representation, then the system $g = Gv$ has a component-wise nonnegative solution $(\pi', \bar{v}')'$.

Now suppose $g = Gv$ has a component-wise nonnegative solution $(\pi', \bar{v}')'$, we want to show that this π is a set-monotone stable representation of ρ. By the definition of G, π is a complete (i.e., includes all possible menus) collection of distributions over $U \times \Phi$. Moreover, the constructed π is set-monotone and stable and can generate the observed ρ.

A.3. Proof of Proposition 2

Towards a contradiction assume $\Pi_\rho = \Delta(U)$. If ρ is irregular, then there exist $A, B \in \mathcal{A}$, $A \subseteq B$, and $a \in A$ such that $\epsilon \equiv \rho_B(a) - \rho_A(a) > 0$. Since by assumption $\Pi_\rho = \Delta(U)$, take any π such that a is the worst with probability 1. If ϕ^* is such that $\phi^*(A) = \phi^*(B) = \{a\}$, then

$$\rho_B(a) = \sum_{\succ} \sum_{\phi} \pi_B(\phi | \succ) \pi(\succ) \mathbb{1} \left(a \succ b, \, \forall b \in \phi(B), \, a \in \phi(B) \right) = \sum_{\succ} \pi_B(\phi^* | \succ) \pi(\succ).$$

Similarly,

$$\rho_A(a) = \sum_{\succ} \pi_A(\phi^* | \succ) \pi(\succ).$$

Taking the difference between these two equations we get that

$$0 < \epsilon = \sum_{\succ} [\pi_B(\phi^* | \succ) - \pi_A(\phi^* | \succ)] \pi(\succ) \leq 0,$$

where the last inequality follows from set-monotonicity. This contradictions completes the proof.