

Identification in Stochastic Choice

Pete Caradonna¹ Christopher Turansick²

¹Division of the Humanities and Social Sciences
California Institute of Technology

²Department of Decision Sciences
Bocconi & IGIER

UC Davis
February 3rd, 2026

Fundamental Question: Given model of behavior, to what degree can underlying primitives can be uniquely determined from individual/aggregate data?

- Even when impossible, knowledge of identified set crucial for design of policy, counterfactual simulation, welfare analysis etc.

Introduction

Despite ubiquity, identification properties of many work-horse models of stochastic choice remain poorly understood.

E.g. for the **random utility model** (with menu variation):

- (i) Known to be unidentified (Fishburn 1998).
- (ii) Non-identification is generic (McClellon 2015).
- (iii) Characterization of *which* data admit unique rationalization (Turansick 2022).

In response, extensive literature has emerged studying additional restrictions which allow for (unique) inversion of choice shares.

An Example

When analyst can't observe individual-level identifying data, population distribution over preferences can't, in general, be uniquely recovered from choice shares.

Classic Example: Due to Fishburn (1998)

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
a	b	a	b
b	a	b	a
c	d	d	c
d	c	c	d

Similar difficulties arise for many other stochastic choice theories.

This Paper

We consider an empiricist who is able to observe choice frequency data across variety of decision problems.

Basic Question: Given a stochastic choice model, and consistent set of observations, what is the *set* of distributions over choice types consistent with the data? Which restrictions on a model are *identifying*?

- Focus on finite, discrete environments.
- Novel, underlying mathematical result on decompositions of network flows; provides complete answers to the above for variety of different stochastic choice models and settings.

Why Care About Identified Sets?

- Parameters/'behavioral types' of models often of direct economic interest, beyond choice shares alone.
 - E.g. distributions of time/risk attitudes, attention filters, choice frames, etc. in society.
- Effects of policy intervention often best understood conditional on underlying preference (or more general choice behavior).
 - Bounds on population shares of different preferences yield bounds on estimated policy impact.
- Identification failures arise from inability of choice data to pin down *correlation* across choices. Particular choices, in turn, may be correlated with demographic features.
 - Implications for, e.g., distributional concerns in policy/welfare analysis.

This Talk: The Random Utility Model

Random Utility Model: Finite set of alternatives X , observe stochastic choice rule:

$$\rho(a, A) = \mathbb{P}(a \text{ chosen from } A),$$

for each $a \in A \subseteq X$. Rationalization is prob. measure μ over linear orders $\mathcal{L}$ such that:

$$\rho(a, A) = \mu\{\succ : \succ \text{ ranks } a \text{ maximally in } A\},$$

for all $a \in A \subseteq X$.

→ **Goal**: Characterize the set of distributions in $\Delta(\mathcal{L})$ consistent with a given, rationalizable ρ .

Choice Functions

Definition

A **choice function** on X is map $\mathbf{c} : 2^X \setminus \{\emptyset\} \rightarrow X$ such that $\mathbf{c}(A) \in A$ for all $\emptyset \subsetneq A \subseteq X$.

Denote set of all choice functions $\mathcal{C}$, and $\mathcal{C}_R \subseteq \mathcal{C}$ set of rational choice functions, i.e.

$$\mathcal{C}_R = \{\mathbf{c}_{\succ} \in \mathcal{C} : \succ \in \mathcal{L}\},$$

where $\mathbf{c}_{\succ}$ denotes constrained-optimal choice according to $\succ$.

Correlation as Source of Identification Failure

By definition:

$$\mathcal{C} = \prod_{\emptyset \subsetneq A \subseteq X} A.$$

Thus:

- (i) A measure $\mu \in \Delta(\mathcal{C})$ is compatible with choice frequencies ρ if, for each $\emptyset \subsetneq A \subseteq X$, the p.m.f. $\rho(\cdot, A)$ coincides with marginal of μ over A .
- (ii) $\mu \in \Delta(\mathcal{C})$ is a RUM-rationalization if, additionally, $\text{supp}(\mu) \subseteq \mathcal{C}_R$.

Takeaway: Want to characterize sets of (i) measures with prescribed marginals, subject to (ii) side constraints on support, corresponding to rationality restrictions.

A Visual Example

Example

Suppose $X = \{a, b, c\}$ and choice frequencies are observed only for menus $\{a, b\}$, $\{b, c\}$, and $\{a, c\}$. Set of choice functions $\mathcal{C} = \{a, b\} \times \{b, c\} \times \{a, c\}$.

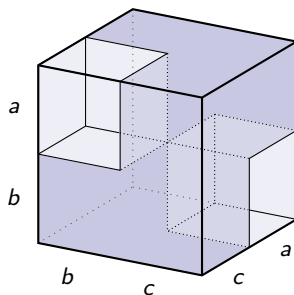


Figure: Set of choice functions $\mathcal{C}$, and $\mathcal{C}_R$, the subset of rational choice functions (shaded). The two irrational functions correspond to cycles $a \succ b \succ c \succ a$ (top left) and $a \succ c \succ b \succ a$ (bottom right).

A Visual Example

Example

Suppose, now, we observe uniform choice probabilities $(\frac{1}{2}, \frac{1}{2})$ on each menu. Then there exist multiple, distinct, *uniform* rationalizations!

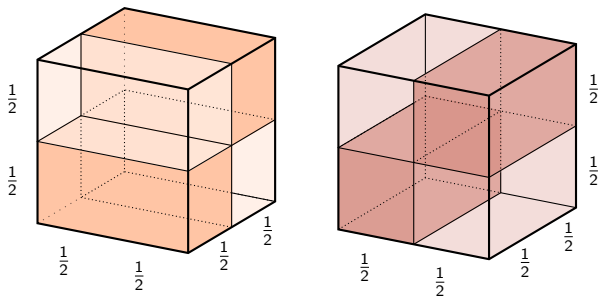


Figure: Under left rationalization, knowledge of subject's choice on $\{a, b\}$ (vertical axis) suffices to perfectly predict choice on $\{a, c\}$ (bottom right axis). Conversely, under right rationalization, knowledge of choice on $\{a, b\}$ is perfectly *uninformative* for choice over $\{a, c\}$.

Modifying Correlation, While Preserving Rationality

Consider again Fishburn's unidentified example:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
a	b	a	b
b	a	b	a
c	d	d	c
d	c	c	d

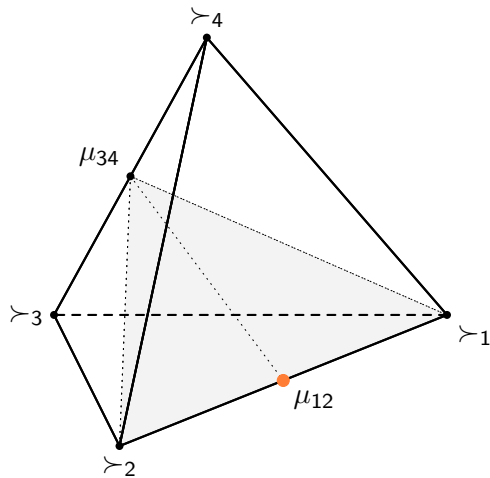
Here, uniform distributions μ_{12} on $\{\succ_1, \succ_2\}$ and μ_{34} on $\{\succ_3, \succ_4\}$ observationally equivalent.

Special structure: $\succ_1$ and $\succ_2$ agree on (i) top two alternatives, but not their ranking, and (ii) bottom two alternatives, but not their ranking.

→ Given any measure μ over these orders, swapping any ε mass from both $\succ_1$ and $\succ_2$ to $\succ_3$ and $\succ_4$ (or vice-versa) preserves all choice probabilities!

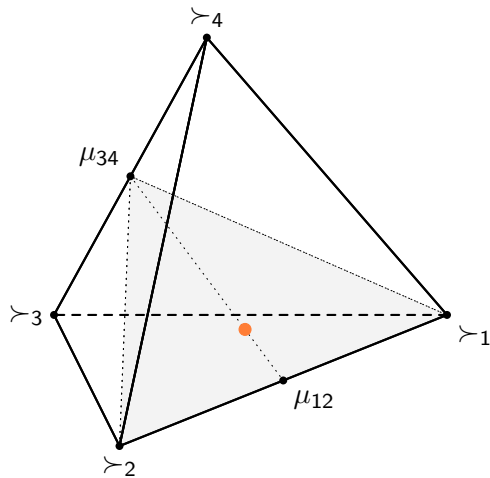
Modifying Correlation, While Preserving Rationality

	cd	dc
ab	$\succ_1$	$\succ_3$
ba	$\succ_4$	$\succ_2$



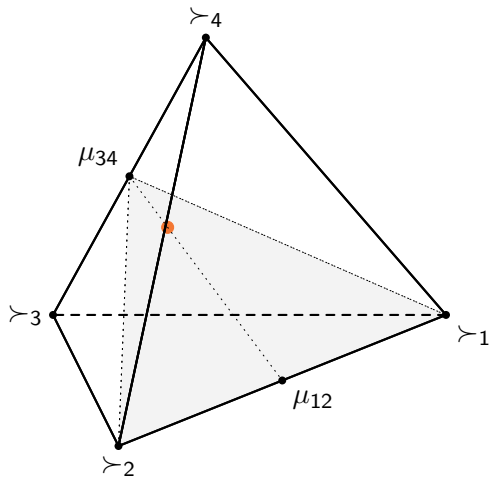
Modifying Correlation, While Preserving Rationality

	cd	dc
ab	$\succ_1$	$\succ_3$
ba	$\succ_4$	$\succ_2$



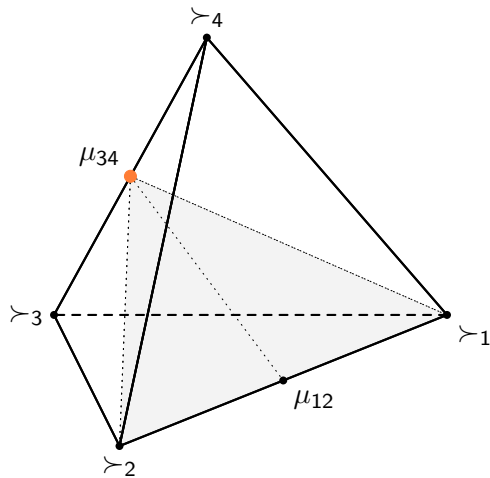
Modifying Correlation, While Preserving Rationality

	cd	dc
ab	$\succ_1$	$\succ_3$
ba	$\succ_4$	$\succ_2$



Modifying Correlation, While Preserving Rationality

	cd	dc
ab	$\succ_1$	$\succ_3$
ba	$\succ_4$	$\succ_2$



Compatible Preferences

Definition

We say two preferences are **k -compatible** if they agree on the set of k best alternatives.

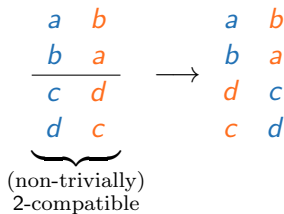
If, in addition, they *disagree* on the ranking of the top k alternatives, and of the remaining $|X| - k$, we say they are **non-trivially k -compatible**.

Intuition: Gets at the same 'product'-like structure as before.

Conjugate Preferences

Definition

The $(k-)$ **conjugate** of a pair of $(k-)$ compatible preferences is the pair obtained by swapping rankings of the $|X| - k$ worst alternatives.



Idea: Compatibility ensures swap yields well-defined preferences. Non-triviality $\leftrightarrow$ conjugates are *distinct* from original pair.

An Observation

Proposition

Let $\succsim_1, \succsim_2$ be compatible, with conjugates $\succsim_3, \succsim_4$. Then the uniform distributions on $\{\succsim_1, \succsim_2\}$ and $\{\succsim_3, \succsim_4\}$ are observationally equivalent.

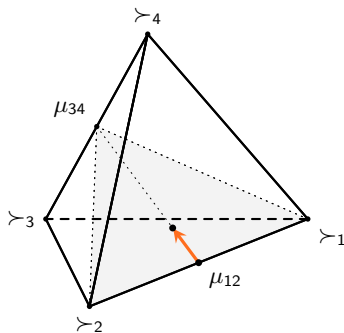
Ryser Swaps

A **weighted Ryser swap** is a (signed) measure in $\mathbb{R}^{\mathcal{L}}$ of the form:

$$\varepsilon \left(\mathbb{1}_{\{\gamma'', \gamma'''\}} - \mathbb{1}_{\{\gamma, \gamma'\}} \right),$$

where $\{\gamma'', \gamma'''\}$ are conjugate to the compatible pair $\{\gamma, \gamma'\}$ for some $k = 1, \dots, |X|$.

Intuition: Adding weighted Ryser swap factors out small product measure, applies marginal-preserving transform, then recombines to generate a new distribution.



A More Complex Example

However, in richer environments, observationally equivalent distributions need not differ only by weighted Ryser swap.

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$
a	b	c	a	b	c
b	a	d	b	a	d
c	e	b	e	c	b
d	f	a	f	d	a
e	c	f	c	f	e
f	d	e	d	e	f

Here, μ_{123} , the uniform distribution on $\{\succ_1, \succ_2, \succ_3\}$ is behaviorally equivalent to the uniform μ_{456} on $\{\succ_4, \succ_5, \succ_6\}$. But any swap necessarily creates preferences outside $\{\succ_1, \dots, \succ_6\}$!

A More Complex Example

γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
a	b	c	a	b	c
b	a	d	b	a	d
c	e	b	e	c	b
d	f	a	f	d	a
e	c	f	c	f	e
f	d	e	d	e	f

However, μ_{123} and μ_{456} still differ by a *sequence* of such swaps.

a	b	c		a	b	c		a	b	c
b	a	d		b	a	d		b	a	d
c	e	b		e	c	b		e	c	b
d	f	a	→	f	d	a	→	f	d	a
e	c	f		c	e	f		c	f	e
f	d	e		d	f	e		d	e	f

The Geometry of Identification

Theorem

Two distributions over preferences are behaviorally equivalent if, and only if, they can be obtained from one other by a finite sequence of weighted Ryser swaps.

Let $\mathcal{R} \subseteq \mathbb{R}^{\mathcal{L}}$ denote the **Ryser subspace**, the span of the weighted Ryser swaps.

Corollary

The set of distributions behaviorally equivalent to $\mu \in \Delta(\mathcal{L})$ are precisely:

$$(\mu + \mathcal{R}) \cap \Delta(\mathcal{L}).$$

Takeaway: Not only do weighted Ryser swaps preserve choice probabilities, they generate the collection of *all* such transformations.

Identifying Restrictions

Definition

A **restriction** on the random utility model is any subset of measures $\mathcal{M} \subseteq \Delta(\mathcal{L})$.

Intuition: We think of $\mathcal{M}$ as representing ‘permissible’ distributions of types in the population.

→ Immediate characterization of identifying restrictions: $\mathcal{M}$ is identifying if and only if, for every $\mu \in \mathcal{M}$,

$$(\mu + \mathcal{R}) \cap \mathcal{M} = \{\mu\}.$$

Often times, however, interested in $\mathcal{M}$ with particular form, in which case, more can be said.

Support Restrictions

Practical Consideration: An analyst may wish to a priori restrict the *set* of preferences assumed held by a population, but not their relative frequencies.

- EU prefs over lotteries (Gul & Pesendorfer '06, Apesteguia & Ballester '18)
- SEU over acts (Lu '16)
- Exponential/quasi-hyperbolic discounting over streams (Lu & Saito '18)
- Single-peaked preferences on ordered domains (Apesteguia et al. '17)

Various separability/forward-looking type assumptions in dynamic models (e.g. Rust '87, Frick et al. '19).

A Simple Example

Suppose $X = \{a, b, c, d\}$ consists of the following monetary lotteries:

$$\begin{aligned}a &= \frac{1}{2}[\$10] + \frac{1}{2}[\$20], & b &= \frac{1}{2}[\$5] + \frac{1}{2}[\$25] \\c &= \frac{1}{4}[\$10] + \frac{3}{4}[\$20], & d &= \frac{1}{4}[\$5] + \frac{3}{4}[\$25].\end{aligned}$$

Wish to consider only (restrictions of) monotone, EU preferences. Note $c \geq_{sd} a$ and $d \geq_{sd} a$.
Moreover:

$$c = \frac{1}{2}[a] + \frac{1}{2}[\$20] \quad \text{and} \quad d = \frac{1}{2}[b] + \frac{1}{2}[\$25],$$

thus $b \succ a$ implies $d \succ c$. This yields:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$
d	d	d	c	c
c	c	b	d	a
b	a	c	a	d
a	b	a	b	b

Support Restrictions

Definition

We say $\mathcal{M}$ is a **support restriction** if $\mathcal{M}$ equals the face of $\Delta(\mathcal{L})$ spanned by some sub-collection of orders $\mathcal{S} \subseteq \mathcal{L}$.

For example, which measures in $\Delta(\succ_1, \dots, \succ_6)$ are behaviorally equivalent to the uniform distribution?

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$
a	b	c	a	b	c
b	a	d	b	a	d
c	e	b	e	c	b
d	f	a	f	d	a
e	c	f	c	f	e
f	d	e	d	e	f

Here, any single weighted Ryser swap supported on preference(s) outside $\{\succ_1, \dots, \succ_6\}$.

Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'

a	b	c		a	b	c
b	a	d				
e	c	b				
f	d	a	→			
c	f	e				
d	e	f				

Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'

a	b	c		a	b	c
b	a	d		b	a	d
e	c	b	→			
f	d	a				
c	f	e				
d	e	f				

Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'

a	b	c		a	b	c
b	a	d		b	a	d
e	c	b		c	e	b
f	d	a	→			
c	f	e				
d	e	f				

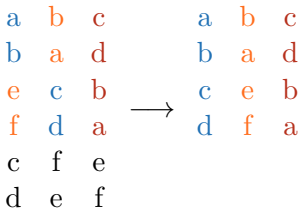
Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'



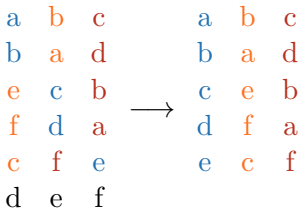
Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'



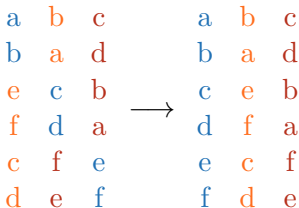
Rearrangements

Let $\succ_1, \dots, \succ_N$ be a sequence of preferences (allowing repetition) drawn from $\mathcal{S}$.

For each $2 \leq k \leq |X|$, obtain a partition Π_k corresponding to the sets of indices of $(k-1)$ -compatible preferences.

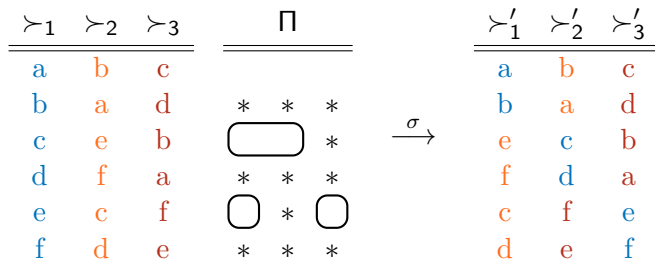
A **rearrangement** $\{\sigma_k\}_{k=2}^{|X|}$ is a collection of permutations of $\{1, \dots, N\}$ such that σ_k fixes each partition element of Π_k .

→ Induces new sequence of preferences by carrying out all these re-assignments from the 'top down.'



Rearrangement Equivalence

We say two sequences are **rearrangement-equivalent** if there exists a rearrangement that transforms one into the other.



Intuition: Visually, writing preferences as strings, rearrangement equivalence $\leftrightarrow$ can “braid” the strings of one sequence to form the other.

A Dual Characterization of Identified Sets

Theorem

Suppose $\mathcal{M}$ is a support restriction. Then $\mu \in \mathcal{M} \subseteq \Delta(\mathcal{L})$ is an extreme point of its $\mathcal{M}$ -identified set:

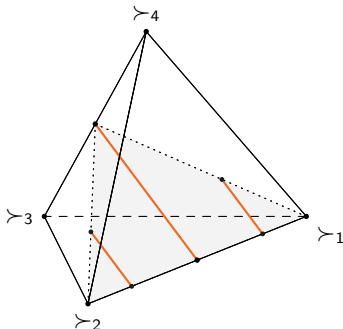
$$(\mu + \mathcal{R}) \cap \mathcal{M}$$

if and only if $\text{supp}(\mu)$ contains no pair of distinct, rearrangement-equivalent sequences.

Fishburn, Revisited

Suppose $\mathcal{S} = \{\succ_1, \dots, \succ_4\}$, where:

γ_1	γ_2	γ_3	γ_4
a	b	a	b
b	a	b	a
c	d	d	c
d	c	c	d



Any distribution with full support is not extremal, as sequences $\succ_1, \succ_2$ and $\succ_3, \succ_4$ are related by a rearrangement (where only σ_2 is non-trivial). However:

- No proper subset of $\mathcal{S}$ contains pair of distinct, rearrangement-equivalent sequences.
- Here, $\mu \in \mathcal{M}$ is an extreme point of its $\mathcal{M}$ -identified set if and only if $\text{supp}(\mu) \subsetneq \mathcal{S}$.

Identifying Support Restrictions

Corollary

Suppose $\mathcal{M}$ is a support restriction corresponding to some $\mathcal{S} \subseteq \mathcal{L}$. Then the following are equivalent:

- (i) $\mathcal{M}$ is identifying.
- (ii) For every $\mu \in \Delta(\mathcal{L})$ supported on $\mathcal{S}$, μ is the unique measure in $(\mu + \mathcal{R}) \cap \Delta(\mathcal{L})$ with this property.
- (iii) $\mathcal{S}$ contains no pair of distinct, rearrangement-equivalent sequences.

Expected Utility, Revisited

Recall that in our four-lottery example, we obtained:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$
d	d	d	c	c
c	c	b	d	a
b	a	c	a	d
a	b	a	b	b

as the strict preferences arising as restrictions of monotone expected utility functions.

- Here, preferences can only be non-trivially k -compatible if $k = 2$; only σ_2 matters.
- To generate new prefs, only option is to swap ab and ba between prefs which differ in rankings of dc and cd ... but can't do so while remaining in $\mathcal{S}$.

Bottom Line: Restriction to EU preferences is identifying on $X = \{a, b, c, d\}$.

Ordered Environments

In many settings, there is some natural ordering $\succeq$ over alternatives X .

- E.g. lotteries by various risk measures/stochastic dominance, vehicles by emission levels, government policies by level of public good provision etc.

Gives rise to an ordering over preferences (or more generally, choice functions).

- One preference dominates another if it is 'more aligned' with $\succeq$, in various senses. Often partial/incomplete.

Many recent identification results involve order-based restrictions. Provide unique selection from identified set, and *linearly* ordered support somehow reflecting $\succeq$.

- Linear ordering of support is useful for statics, provides 'one-dimensional' type space for mechanism design etc.

Order-Based Restrictions

However, existing order-based restrictions all involve trade-offs. Fix linear order $\succeq$ of alternatives.

Some prominent examples:

- **Single-Crossing Random Utility** (Apesteguia et al. '17): Require support of rationalizing $\mu \in \Delta(\mathcal{L})$ to satisfy single-crossing property with respect to $\succeq$.
 - More restrictive than random utility: exist RUM- but not SCRUM-rationalizable ρ .
- **Progressive Random Choice** (Filiz-Ozbay & Masatlioglu '23): Choose unique rationalization $\mu \in \Delta(\mathcal{C})$ whose support is a chain in the product order on $\mathcal{C}$.
 - Impossible to choose order $\succeq$ to guarantee PRC rationalization is supported on $\mathcal{C}_R$ (i.e. $\mathcal{L}$) whenever ρ is RUM-rationalizable.

Swap-Progressivity

Fix any linear ordering of alternatives $\triangleright$.

Definition

A measure $\mu \in \Delta(\mathcal{L})$ is said to be **swap-progressive** if $\text{supp}(\mu)$ can be linearly ordered such that, for any $\succ_i, \succ_j \in \text{supp}(\mu)$ which are k -compatible for some $0 \leq k \leq |X| - 1$:

$$i > j \quad \implies \quad x_i^{k+1} \triangleright x_j^{k+1}$$

where x_i^{k+1} (resp. x_j^{k+1}) denotes the $(k+1)$ -st highest ranked alternative under $\succ_i$ (resp. $\succ_j$).

Takeaway: Novel restriction, neither implies nor implied by progressivity or single-crossing. Only impose requirements for pairs of *compatible* preferences.

Swap-Progressivity: An Example

Let $X = \{a, b, c, d\}$ be government policies, ordered by their fiscal cost $a \succeq b \succeq d \succeq c$, and suppose data ρ are consistent with uniform distribution on:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
a	b	a	b
b	a	b	a
c	d	d	c
d	c	c	d

This distribution cannot be swap progressive. Indeed no swap-progressive rationalization can put positive mass on both $\succ_1$ and $\succ_2$:

- As $\succ_1$ and $\succ_2$ are (trivially) 0-compatible, $a \succeq b$ implies $\succ_1$ would have to be ranked higher in the support ordering than $\succ_2$.
- But $\succ_1$ and $\succ_2$ are 2-compatible; as $d \succeq c$, $\succ_2$ would simultaneously have to be ranked higher than $\succ_1$, an impossibility.

Swap-Progressivity: Best of Both Worlds

Theorem

Let $\succeq$ be a linear order on X . Then:

- (i) A swap-progressive rationalization exists if and only if the data are compatible with the random utility model. Moreover, if such a representation exists, it is unique.
- (ii) For any $\succeq$, a single-crossing random utility representation exists if and only if a progressive one does, and in this case both these and the unique swap-progressive rationalization all coincide.

Takeaway: Principled refinement of both single-crossing and progressive frameworks. Agrees with random utility predictions of each, when these exist, but generates unique predictions even when these fail.

Other Properties of Swap-Progressive Representations

Proposition

Let $\succeq$ be a linear order on X . Then:

- (i) The unique swap-progressive (hence single-crossing/progressive, when these exist) random utility rationalization is always an extreme point of its identified set.
- (ii) The support ordering of the swap-progressive rationalization is always a restriction of the dictionary order on preferences (regarded as strings) induced by $\succeq$.

Parametric Restrictions

In practical applications, random utility, and more general models of stochastic choice, often specified parametrically.

If $\Theta \subseteq \mathbb{R}^d$ denotes a set of parameters, we refer to $F : \Theta \rightarrow \Delta(\mathcal{L})$ as a **parametric random utility model**. In principle, such a model fails to be identified if and only if:

$$F(\theta') \in F(\theta) + \mathcal{R}, \quad (1)$$

for distinct $\theta, \theta' \in \Theta$, i.e. if $F(\theta)$ and $F(\theta')$ belong to some common translate of the Ryser subspace.

Difficulty: In practice, not clear how to verify this from inspection of F . Moreover, choice probabilities often specified directly (e.g. Luce), leaving F only implicit.

Parametric Restrictions

We take a general approach—no longer restrict to only random utility.

Let $\mathcal{P}$ denote set of random choice rules. Suppose $\mathcal{M} \subseteq \mathcal{P}$ is smoothly parameterized by some function φ , i.e.

$$\varphi : \Theta \rightarrow \mathcal{M} \subseteq \mathcal{P}$$

is a smooth mapping of some convex, open set of parameters $\Theta \subseteq \mathbb{R}^d$.

Definition

Say φ is **parametrically identified** if there exists some smooth inverse $\varphi(\Theta) \rightarrow \Theta$.

Question: Can we give simple test for parametric identification of φ that doesn't require intimate geometric understanding of $\mathcal{M}$?

Parametric Submodels

Problem: In general, it's intractable to directly analyze full vector of functions φ in any realistically-sized setting.

→ If $|X| = 6$, general random choice rule 129-dimensional.

To determine invertibility in practice, typically restrict to subsystems with equal numbers of parameters and choice probabilities.

Definition

A smooth map $\bar{\varphi} : \Theta \rightarrow \mathbb{R}^d$ is a **submodel** if it's obtained by selecting d components of the vector-valued map $G \circ \varphi$, where G is any smooth and smoothly-invertible map.

Example: Luce Model

Let $X = \{x_0, \dots, x_K\}$. A random choice rule ρ belongs to the **Luce model** if there exist weights $w_0 \equiv 1$ and $w_k > 0$ for all $k > 0$, such that, for any $x \in A \subseteq X$:

$$\varphi(w_1, \dots, w_K)_{(x_i, A)} = \rho(x_i, A) = \frac{w_i}{\sum_{j \in A} w_j}.$$

Thus $\Theta = \mathbb{R}_{++}^K$, with each θ corresponding to the vector of normalized weights $(w_1, \dots, w_K)$. Consider the map:

$$\bar{\varphi}(w_1, \dots, w_K) = \begin{bmatrix} \frac{\rho(x_1, X)}{\rho(x_0, X)} \\ \vdots \\ \frac{\rho(x_K, X)}{\rho(x_0, X)} \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_K \end{bmatrix}.$$

Here G normalizes each component of φ by $\rho(x_0, X)$; $\bar{\varphi}$ then just selects the $d = K$ components corresponding to normalized choice probabilities for $x_1, \dots, x_K$ from X .

Parametric Identification of Submodels

Theorem

Let $\Theta \subseteq \mathbb{R}^d$ be a convex, open set of parameters. A smooth map $\bar{\varphi} : \Theta \rightarrow \mathbb{R}^d$ is parametrically identified if and only if:

- (i) The Jacobian matrix $\mathcal{J}_{\bar{\varphi}}$ is everywhere invertible; and
- (ii) For any sequence $(\theta_n)_{n=1}^{\infty} \subset \Theta$ such that each compact $K \subset \Theta$ contains at most finitely many terms, the same is true of $(\bar{\varphi}(\theta_n))_{n=1}^{\infty}$ for any compact $K' \subset \bar{\varphi}(\Theta)$.

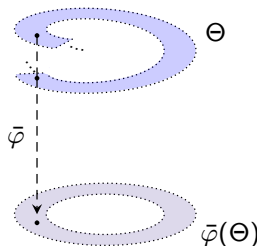
A Few Remarks

- Our inversion theorem applies only to maps with the same number of variables as equations.
 - While same techniques can treat general case, leads to additional, difficult-to-verify requirements on range of φ .
 - In practice, these typically checked *precisely by showing a submodel is parametrically identified*. Hence we emphasize the submodel formulation.
 - Parametric identification of $\bar{\varphi}$ implies the same of φ . Converse not generally true, but frequently valid in practice. See paper for illustration.
- Our proof purely topological. This allows us to avoid common, monotonicity-type assumptions which tend to be sufficient but not necessary.
 - E.g. applied to demand invertibility questions, our result applies even when goods may not be substitutes, or when the law of demand fails to hold.

Necessity of The 'Boundary Condition'

Informally, condition (ii) said that $\bar{\varphi}$ maps sequences tending to the boundary of Θ to sequences which tend to the boundary of $\text{range}(\bar{\varphi})$.

Unfortunately, it cannot be dropped.



Similar issues crop up, e.g., in Gale-Nikaido '65.

Luce, Revisited

Recall we defined $\Theta = \mathbb{R}_{++}^K$, and:

$$\bar{\varphi}(w_1, \dots, w_N) = \begin{bmatrix} w_1 \\ \vdots \\ w_K \end{bmatrix}.$$

The Jacobian $\mathcal{J}_{\bar{\varphi}}$ is everywhere the K -dimensional identity matrix, and trivially if some component of (θ_n) goes to zero or infinity, so does $\bar{\varphi}(\theta_n) = \theta_n$. Thus $\bar{\varphi}$ (and hence the Luce model itself) are parametrically identified.

A More Complex Example

DLHF Model: Suppose $X = \{x_0, \dots, x_N\}$, where x_0 denotes some outside option.

Random choice rule represents the *unconditional* choice probabilities of the dynamic logit habit formation model of Turansick (2025) if, for any menu $x_0 \in A \subseteq X$:

$$\varphi(v_1, c_1, \dots, v_N, c_N)_{(x_i, A)} = \rho(x_i, A) = \frac{v_i \left(\sum_{j \in A \setminus i} v_j + v_i c_i \right)}{\sum_{j \in A} v_j \left(\sum_{k \in A \setminus j} v_k + v_j c_j \right)},$$

for $v_i, c_i > 1$, where $v_0 = c_0 \equiv 1$. Here $\Theta = (1, \infty)^{2N}$, i.e. $\theta = (v_1, c_1, \dots, v_N, c_N)$.

Reduction to Submodel

Define, for $i = 1, \dots, N$:

$$\bar{\varphi}^i(v_1, c_1, \dots, v_N, c_N) = \frac{\rho(x_0, \{x_0, x_i\})}{\rho(x_i, \{x_0, x_i\})} = \frac{1 + v_i}{v_i(1 + v_i c_i)}$$

and

$$\bar{\varphi}^{N+i}(v_1, c_1, \dots, v_N, c_N) = \frac{\rho(x_0, \{x_0, x_j, x_{j+1}\})}{\rho(x_i, \{x_0, x_i, x_{i+1}\})} = \frac{1 + v_i + v_{i+1}}{v_i + v_i v_{i+1} + c_j v_i^2},$$

where $v_{N+1} \equiv v_1$, and $c_{N+1} \equiv c_1$.

→ Given data, non-linear system with equal numbers of equations ($2N$) as unknowns ($2N$).

Failure of Parametric Identification of $\bar{\varphi}$

Define the function:

$$c(v) = \frac{1+v}{10v^2} - \frac{1}{v}.$$

Chosen so, for $i = 1, \dots, N$:

$$\bar{\varphi}(v_1, c(v_1), \dots, v_N, c(v_N)) = \text{const.}$$

But, along the curve $(v, c(v), \dots, v, c(v))$,

$$\bar{\varphi}^{N+i}(v, c(v), \dots, v, c(v)) = \frac{1+2v}{v^2 + 10(1+v)}.$$

Limit of this as $v \rightarrow 1$ is $\frac{1}{7}$. But this is same as value at $v = 3$. Hence condition (ii) of parametric identification theorem fails: $\bar{\varphi}$ not identified.

→ **In Paper:** simple to use this observation to show non-identification of φ as well.

Beyond Random Utility

- Each theorem today has counterparts for a variety of other stochastic choice applications.
 - General random choice distributions, under possibly incomplete data.
 - Choice histories from conditional choice frequencies in dynamic models.
 - Type distributions in boundedly rational stochastic choice theories, such as random attention, framing, etc.
 - Cross-agent choice correlations from joint choice data in multi-agent models.
- Common underlying structure is that these problems can all be encoded as problems about decomposing flows on particular directed graphs.
 - Connections to mathematical literature on tomography and measures with prescribed marginals.

This Paper

Characterizing identification requires understanding possible correlations between 'segments' of preferences that leave choice probabilities unaffected.

- Introduce notion of Ryser swap, which provides simple description of identified sets.
- Unifying tool for studying identification properties of many common forms of restriction (e.g. support, order-based, etc.)
 - Yields practical insights: novel test for identifying support restrictions, new identifying order-based restrictions (swap-progressivity).
- Tractable, powerful test for parametric identification.

In the Paper: Notion of Ryser swaps for various dynamic, boundedly rational, and multi-agent models. Parallel results for these settings.

Conclusions

- When analysts observe only aggregate choice frequency data, many work-horse models of stochastic choice are unidentified.
- Explicit descriptions of identified sets of direct practical importance: allow for not only understanding range of possible aggregate effects of policies, but also distributional implications.
- Provide first characterizations of these sets, for wide range of theories, via unified approach. Representation in terms of novel 'swapping' transforms we term *Ryser swaps*. Provides tools for practitioners to investigate identifying restrictions.

Thank You!

Any Questions?