ECO499 Proposal: Revised Edition

Andrew Scutt[1]

[1]Faculty of Economics
University of Toronto

Proposal, October 2025

Table of Contents

Table of Contents

Knowledge Landscape

As a preface, this proposal focuses on modeling discrete choices of a population using a characteristic-based approach.

In the early 20th century, choice modeling research was restricted to the binomial choice framework: researchers had to limit their inquiries to binary choices between two items. Afterward, they had to infer the overall preference distribution using this set of pairwise comparisons. A prime example would be Thurstone's paper for estimating the importance of certain social values (Thurstone, 1927). Unfortunately, this approach was extremely time-consuming due to the number of comparisons required, being $\binom{n}{2} = \frac{n!}{2!(n-2)!}$, where n is the number of items. In addition, this method allowed intransitive preferences to seep into the overall preference data due to its reliance on raw pairwise comparisons. To account for preferences involving more than two choices, statisticians developed the conditional logistic model, which circumvents intransitivity problems. It allows us to model the choice as a function of item characteristics.

Knowledge Landscape

For the following discussion, let

$$\arg\max_{k \in K} \Pr(Y = k \mid X = x)$$

denote the choice most likely selected by the participant, where K indexes the available bundles. The choice set is

$$K = \{k_1, k_2, \dots, k_m\},$$

where each

$$k_j = (x_{j,1}, x_{j,2}, \dots, x_{j,n}) \in \mathbb{R}^n$$

is an n-dimensional bundle. Conditional logit models characterize the log-odds of choosing bundle j over a baseline bundle m as a linear function of attribute differences:

$$\log\left(\frac{P_{i,j}}{P_{i,m}}\right) = \beta_1(x_{i,j,1} - x_{i,m,1}) + \beta_2(x_{i,j,2} - x_{i,m,2}) + \cdots + \beta_p(x_{i,j,n} - x_{i,m,n}),$$

Knowledge Landscape

where $P_{i,j}$ is the probability that participant i selects bundle j. Simply put, this model is trained on a set of options coupled with each participant's choices (yes/no). Then, the user inputs a new set of items to which the algorithm outputs the probability of choosing each item. However, this model assumes a linear relationship across attributes, limiting their ability to capture truly nonlinear effects; our method can overcome this issue. In addition, our statistical approach can extrapolate beyond the menu size structure used for training this model.

Table of Contents

Setup

Let X be the finite set of all possible alternatives that our agent might be facing. Let $\mathcal{A} = 2^X \backslash \{\varnothing\}$ be the collection of all non-empty and finite subsets of X, with typical elements A, B, and C, which we call menus. Let $\psi = \bigcup_{A \in \mathcal{A}} \text{Perm}(A)$ be the collection of all nonempty and finite strict preference orderings of $\mathcal{A}$. We define a ranking function $\mathcal{X}$: $\mathcal{A} \rightarrow \psi$, such that $\mathcal{X}(A) \in \text{Perm}(A)$. A binary relation $\succ$ on X is a called a strict preference if it is complete $(\forall x, y \in X : \; x \succeq y \;\; \vee \;\; y \succeq x)$, transitive $(\forall x, y, z \in X : \; (x \succeq y \; \wedge \; y \succeq z) \; \Rightarrow \; x \succeq z)$, and antisymmetric $(\forall x, y \in X : \; (x \succeq y \; \wedge \; y \succeq x) \; \Rightarrow \; x = y)$. We say that a strict preference $\succ$ represents $\mathcal{X}$ whenever $\mathcal{X}(\mathcal{A})$ is ordered according to $\succ$.

Setup

A specific element of $\mathcal{X}(A)$ can be deconstructed into a set of binary strict preferences. Let $\mathcal{B}$: $\psi \to 2^{X \times X}$ be the partial strict preference mapping, $\mathcal{B}((x_1, x_2, \cdots, x_n)) = \{(x_i, x_j)\colon x_i \succ x_j\}$, such that $\mathcal{B}(\mathcal{X}(A)) \in 2^{A \times A}$. (Note: $\mathcal{X}(A) \in \psi$). Now, let $\mathcal{D}$: $\psi \to 2^{X \times X}$ be the complete strict preference mapping, $\mathcal{D}(\text{Perm}(A)) = \bigcup_{d \in \text{Perm}(A)} \mathcal{B}(d)$, such that $\mathcal{D}(\text{Perm}(A)) \in 2^{A \times A}$. (Note: $\text{Perm}(A) \subseteq \psi$)

Global Menu Independence Assumption: $\exists \succ$ on X, such that $\mathcal{X}(A)$ satisfies $\succ|_A \; = \; \{\, (x_i, x_j) \in \mathcal{D}(\text{Perm}(A)) : x_i \succ x_j \text{ in } X \,\}, \forall A \in \mathcal{A}$. In other words, the agent maximizes the same strict preferences on $\mathcal{A}$ irrespective of the menu that they are facing.

Setup

Theorem 1 (Context Independence): If $A \subseteq B$, then $\mathcal{B}(\mathcal{X}(A)) \subseteq \mathcal{B}(\mathcal{X}(B))$.
Proof: Fix $\succ$ on X, such that it satisfies the global menu-independence assumption. Suppose $(x, y) \in \mathcal{B}(\mathcal{X}(A))$. Then $x, y \in A$ and $x \succ y$ w.r.t. the global relation $\succ$. Since $A \subseteq B$, it follows that $x, y \in B$. By menu independence, $\mathcal{X}(B)$ is the restriction of $\succ$ to B, so $x \succ y$ also holds in $\mathcal{X}(B)$. Thus $(x, y) \in \mathcal{B}(\mathcal{X}(B))$. Therefore $\mathcal{B}(\mathcal{X}(A)) \subseteq \mathcal{B}(\mathcal{X}(B))$. $\square$

A preference $\succ$ is represented by a utility function $U : X \to \mathbb{R}$ whenever $x \succ y \iff U(x)U(y)$. If $\succ$ is complete, transitive, and antisymmetric, and X is finite, then $\exists$ a utility representation of $\succ$ always exists.

Ordinal Uniqueness: Functions U_1, U_2 represent the same preference $\succeq$ on $X \iff \exists$ a strictly increasing function $\phi : R_1 \to \mathbb{R}$, such that $U_2(x) = \phi(U_1(x)) \forall x \in X . R_1$ is the range of U_1 defined by $\{U_1(x) : x \in X\}$

Setup

For every $A \in \mathcal{A}$ and $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A))$, let $\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A)))$ denote the frequency with which the ordering (x_i, x_j) was observed.

For any finite set Z, let $\Delta(Z)$ denote the set of probability distributions over Z, functions $\rho : Z \to [0,1]$, such that $\sum_{z \in Z} \rho(z) = 1$. For each menu A, the values of $\rho(., A)$ form a probability distribution over A.

Stochastic Choice Function: *A stochastic choice function* is a mapping $p : \mathcal{A} \to \Delta(X \times X)$, such that
$\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A))) + \rho((x_j, x_i), \mathcal{D}(\text{Perm}(A))) = 1$
$\forall (x_i, x_j), (x_j, x_i) \in \mathcal{D}(\text{Perm}(A))$ and $\forall A \in \mathcal{A}$, and
$\sum_{(x_i, x_j) \in \mathcal{D}(\text{Perm}(A))} \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A))) = \frac{\binom{n}{2}}{2}$ for all $A \in \mathcal{A}$.

To study population-level preferences and account for the notion that there exists an informational asymmetry between our analyst and our agent, we will use random utility.

Setup

Axioms of ρ:

- Regularity: If $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A)) \subseteq \mathcal{D}(\text{Perm}(B))$, then $\rho((x_i, x_j), \mathcal{D}(\text{Perm}(B))) \leq \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A)))$
- Supermodularity: If $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A)) \cap \mathcal{D}(\text{Perm}(B))$, then $\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A))) + \rho((x_i, x_j), \mathcal{D}(\text{Perm}(B))) \leq \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A \cup B))) + \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A \cap B)))$
- Axiom of Revealed Stochastic Preference: For any k and for any sequence $((x_i, x_j), \mathcal{D}(\text{Perm}(A_1))), ..., ((x_i, x_j), \mathcal{D}(\text{Perm}(A_k)))$ for all $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A_l))$ such that $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A_l))$ $\sum_{i=1}^{k} \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A_l)) \leq \max_{\succ \in P} \sum_{i=1}^{k} p_{\succ}((x_i, x_j), \mathcal{D}(\text{Perm}(A_l)).$

- Coherency: For any k and for any sequence $((x_i, x_j), \mathcal{D}(\text{Perm}(A_1))), \ldots, ((x_i, x_j), \mathcal{D}(\text{Perm}(A_k)))$ for all $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A_l))$ such that $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A_l))$ and for any sequence of real numbers $\lambda_1, \ldots, \lambda_k, \sum_{i=1}^{k} \lambda_i p((x_i, x_j), \mathcal{D}(\text{Perm}(A_l))) \geq 0 \implies \sum_{i=1}^{k} \lambda_i \rho((x_i, x_j), \mathcal{D}(\text{Perm}(A_l))) \geq 0.$

Let $\mathcal{P}$ be the set of all strict preferences over a finite X. Let $\mu \in \Delta(\mathcal{P})$ be a probability distribution over these strict preferences. For any $A \in \mathcal{A}$ and $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A))$, let $N((x_i, x_j), \mathcal{D}(\text{Perm}(A))) := \{\succ \in \mathcal{P} : x_i \succ x_j\}$ be the strict preference that rationalizes the choice of (x_i, x_j) from $\mathcal{D}(\text{Perm}(A))$.

$\rho : \mathcal{A} \to \Delta(\mathcal{D}(\text{Perm}(X)))$ is represented by a distribution over preferences if there exists $\mu \in \Delta(\mathcal{P})$, such that $\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A))) = \mu(N((x_i, x_j), \mathcal{D}(\text{Perm}(A))))$ for all $A \in \mathcal{A}$ and $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A))$. Please note that μ must satisfy $\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A)) + p((x_j, x_i), \mathcal{D}(\text{Perm}(A)) = 1$ $\forall (x_i, x_j), (x_j, x_i) \in \mathcal{D}(\text{Perm}(A)$ and $\forall A \in \mathcal{A}$, and $\sum_{x \in \text{Perm}(A)} \rho(x, \text{Perm}(A)) = 1 \; \forall A \in \mathcal{A}$. These restrictions apply to all future discussed probability distributions over $\mathcal{D}(\text{Perm}(X))$.

Setup

Alternative Definitions:
Let $N((x_i, x_j), \mathcal{D}(\text{Perm}(A))) := \{U \in \mathbb{R}^X : U(x_i) > U(x_j)\}$ be the set of strict preferences that rationalize the choice of (x_i, x_j) from $\mathcal{D}(\text{Perm}(A))$ in terms of utility.

Let $\Delta(\mathbb{R}^X)$ be the set of Borel probability measures over $\mathbb{R}^X$.

$\rho : \mathcal{A} \to \Delta(\mathcal{D}(\text{Perm}(X)))$ is represented by a distribution over utilities if there exists $\mu \in \Delta(\mathbb{R}^X)$, such that
$\rho((x_i, x_j), \mathcal{D}(\text{Perm}(A))) = \mu(N((x_i, x_j), \mathcal{D}(\text{Perm}(A))))$ for all $A \in \mathcal{A}$ and $(x_i, x_j) \in \mathcal{D}(\text{Perm}(A))$.

Setup

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, i.e., $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure. Utility is a random function, i.e., $\tilde{U} : \Omega \to \mathbb{R}^X$ is $\mathcal{F}$-measurable.

Let $N((x_i, x_j), \mathcal{D}(\mathrm{Perm}(A))) := \{\omega \in \Omega : \tilde{U}(x_i) > \tilde{U}(x_j)\}$ be the set of strict preferences that rationalize the choice of (x_i, x_j) from $\mathcal{D}(\mathrm{Perm}(A))$ in terms of random utility.

$\rho : \mathcal{A} \to \Delta(\mathcal{D}(\mathrm{Perm}(X)))$ is represented by a random utility if there exists a random variable $\tilde{U} : \Omega \to \mathbb{R}^X$, such that $\rho((x_i, x_j), \mathcal{D}(\mathrm{Perm}(A))) = \mathbb{P}(N((x_i, x_j), \mathcal{D}(\mathrm{Perm}(A))))$ for all $A \in \mathcal{A}$ and $(x_i, x_j) \in \mathcal{D}(\mathrm{Perm}(A))$.

Setup

Now, let's rewrite random utility in terms of additive random utility. Let $\tilde{U}(x) = v(x) + \tilde{\epsilon}(x)$ where $v : X \to \mathbb{R}$ is a deterministic utility function and $\tilde{\epsilon} : \Omega \to \mathbb{R}^X$ is a random utility shock.

If $\rho : \mathcal{A} \to \Delta(\mathcal{D}(\text{Perm}(X)))$ has an additive random utility representation, if it has a RU representation with $\tilde{U}(x) = v(x) + \tilde{\epsilon}(x)$ where $v : X \to \mathbb{R}$ is deterministic and the distribution of $\tilde{\epsilon}$ is smooth.

Setup

For estimating additive random utility, let Θ be the space of parameters and let $(p_\theta)_{\theta \in \Theta}$ be the family of stochastic choice functions indexed by parameter θ. For additive random utility, $\Theta = \mathbb{R}^X \times \Delta(\mathbb{R}^X)$. This is the collection of all deterministic utility functions and the collection of all distributions over shocks.

Setup

Now, let's add attributes to our analysis.

For each alternative x, let ξ_x be a vector of its attributes. Let $\mathcal{E}$ be the space of attributes; assume $\mathcal{E} \subseteq \mathbb{R}^n$ for some n. Let $\mathcal{E}^X$ be the set of all attribute profiles of members of X.

A *s.c.f with attributes* is a function $\rho : \mathcal{E}^X \to \Delta(\mathcal{D}(\text{Perm}(X)))$. We will write $\rho((x_i, x_j), \{\xi_{x_i}, \xi_{x_j}\})$ to mean the probability of choosing x_i over x_j when the vector of i attributes is ξ_{x_i} and the vector of attributes of j is ξ_{x_j}.

Setup

The agent's random utility equals $\tilde{U}(x, \xi_x) = v_x(\xi_x) + \tilde{\epsilon}(x)$

Independent Additive Random Utility: Let X be a finite set.
$\rho : \mathcal{E}^X \to \Delta(\mathcal{D}(\text{Perm}(X)))$ has a simple IARU representation if there exists a random variable $\tilde{\epsilon}$ with values in $\mathbb{R}^X$ distributed smoothly and independently of v, such that
$\rho((x_i, x_j), v) = \mathbb{P}(v(x_i) + \tilde{\epsilon}(x_i) > v(x_j) + \tilde{\epsilon}(x_j))$

For this model, we will fix the distribution of ϵ and not estimate it.

Setup

Now, let's finally move to our proposed model.

Setup

Let $\mathbb{R}^n$ denote the vector space of n-dimensional real-valued vectors.

For each $i \in \{1, \ldots, m\}$, let $s_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n}) \in \mathbb{R}^n$ represent a n-dimensional item.

Let $S = \{s_1, s_2, \ldots, s_m\} \subseteq \mathbb{R}^n$ be the finite set of all such items.

Let there be N participants, indexed by $p \in \{1, \ldots, N\}$, each of whom provides a strict preference ranking for all s_i in S.

Setup

- For each participant p, construct a pairwise comparison matrix $M^p \in \mathbb{R}^{m \times m}$, where:

$$M^p_{i,j} = \begin{cases} 1 & \text{if } s_i \succ_p s_j \\ 0 & \text{if } s_j \succ_p s_i \\ \text{Undefined} & \text{if } i = j \end{cases}$$

  Here, $s_i \succ_p s_j$ indicates that participant p strictly prefers item s_i over s_j, and x denotes an undefined diagonal value.

- Aggregate the individual matrices by summing over all participants:

$$A = \sum_{p=1}^{N} M^p$$

Setup

- Normalize the aggregate matrix by dividing each entry by the number of participants, and add an error term:

$$\bar{A} = \frac{1}{N}A + \tilde{\epsilon}$$

- Interpret the vector $\bar{A} = (\bar{A}_1, \bar{A}_2, \dots, \bar{A}_m)$ as probability that the population would choose each item (to compute $\bar{A}_i$, you need to sum all components in the s_i column) where higher $\bar{A}_j$ indicates stronger preference.
- Afterward, we can represent each item i in the following manner: $(\bar{A}_i, x_{i,1}, x_{i,2}, \dots, x_{i,n})$
- Using this vectorial representation coupled with machine learning techniques, such as symbolic regression, KNN, regression trees, multiple linear regression, etc, we can determine how each dimension $x_{i,s}$ contributes to $\bar{A}_i$ where $s \in \{1, 2, \dots, n\}$. In this case, $\bar{A}_i$ is interpreted as utility.

Setup

Here is an illustration of what the matrix looks like:

$$
\begin{array}{c|cccc}
 & s_1 & s_2 & \cdots & s_n \\
\hline
s_1 & x & \rho((s_2, s_1), v) & \cdots & \rho((s_n, s_1), v) \\
s_2 & \rho((s_1, s_2), v) & x & \cdots & \rho((s_n, s_2), v) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
s_n & \rho((s_1, s_n), v) & \rho((s_2, s_n), v) & \cdots & x
\end{array}
$$

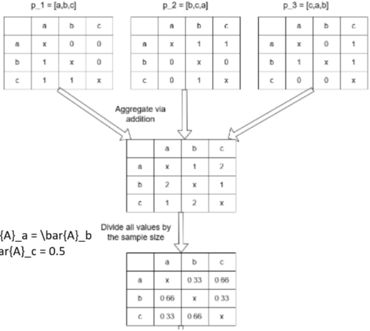
Please note that v is constant across all of its invocations.

Empirical Question

In the context of ECO499, our project will include an empirical innovation by examining how Canadians evaluate multidimensional "society bundles." Each bundle contains attributes such as (i) the presence or absence of a Castle Law–style self-defense provision, (ii) the annual number of murders, (iii) the annual number of robberies, and (iv) the annual number of sexual assaults. Additional dimensions may be added to reflect concerns voiced by both Conservatives and Liberals about potential reasons for and against adopting a Castle Law in Canada.

Drawing on findings from psychology, we hypothesize that individuals with more individualistic cultural values will prefer the presence of a Castle Law even when it is paired with higher crime rates, whereas individuals with more collectivistic cultural values will show the opposite preference. Our research will empirically test this hypothesis by estimating how preferences over these multidimensional bundles vary with respondents' cultural value orientations.

Table of Contents



p_1 = [a,b,c]

	a	b	c
a	x	0	0
b	1	x	0
c	1	1	x

p_2 = [b,c,a]

	a	b	c
a	x	1	1
b	0	x	0
c	0	1	x

p_3 = [c,a,b]

	a	b	c
a	x	0	1
b	1	x	1
c	0	0	x

	a	b	c
a	x	1	2
b	2	x	1
c	1	2	x

	a	b	c
a	x	0.33	0.66
b	0.66	x	0.33
c	0.33	0.66	x

\bar{A}_a = \bar{A}_b = \bar{A}_c = 0.5