Ordered Probabilistic Choice*

Christopher P. Chambers† Yusufcan Masatlioglu‡ Kemal Yildiz§

April 8, 2025

Abstract

We introduce a novel perspective by linking ordered probabilistic choice to copula theory, a mathematical framework for modeling dependencies in multivariate distributions. Each representation of ordered probabilistic choice behavior can be associated with a copula, enabling the analysis of representations through established results from copula theory. We provide functional forms to describe the "extremal" representations of an ordered probabilistic choice behavior–and their distinctive structural properties. The resulting functional forms act as an "identification method" that uniquely generates heterogeneous choice types and their weights. These results provide valuable tools for analysts to identify micro-level behavioral heterogeneity from macro-level observable data.

(JEL: D01, D91)

Keywords: Copulas, choice, correlation, behavioral heterogeneity, Fréchet-Hoeffding bounds

*We thank the audience of BRIC X and the theory seminar participants at Cornell and Caltech.

†Department of Economics, Georgetown University, ICC 580 37th and O Streets NW, Washington DC 20057. E-mail: `Christopher.Chambers@georgetown.edu`.

‡University of Maryland, 3147E Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742. E-mail: `yusufcan@umd.edu`.

§Bilkent University, Department of Economics, Universiteler Mah., Bilkent, 06800, Ankara, Turkey. E-mail: `kemal.yildiz@bilkent.edu.tr`.

Contents

1 Introduction

Ordered probabilistic choice analyzes how individuals make choices among naturally ordered alternatives. Examples abound, including bond ratings, energy consumption, occupational choices, automobile ownership, biological impacts of pesticides, financial investments, and labor force participation rates, among many others. The workhorse model in this area is the random utility model (RUM), which accounts for unobserved heterogeneity in preferences. Although individual choices are deterministic and rational, there is no restriction on conceivable preferences, which makes RUM a highly flexible model. But, this flexibility leads to undesirable identification issues in the form of multiple representations, undermining the interpretability and prediction power of the model.

Apesteguia, Ballester & Lu (2017) solves the challenge of non-unique Random Utility Model (RUM) representations by leveraging the inherent order among alternatives. Their work demonstrates that uniqueness is achieved under three assumptions: i) Ordered alternatives, ii) Rational decision-makers, and iii) Ordered individual preference types. The *single-crossing random utility model* (SCRUM) restricts preference heterogeneity to ordered types while preserving empirical accuracy in capturing individual choices. Expanding on SCRUM, Filiz-Ozbay & Masatlioglu (2023) explored how ordered types apply to boundedly rational agents, incorporating behavioral factors such as limited attention, willpower, shortlisting constraints, loss aversion, and pro-social behavior. While the resulting *progressive random choice* (PRC) model permits a broader spectrum of agent types, it retains the ordered-type framework to accommodate bounded rationality. PRC allows boundedly rational agents without sacrificing uniqueness.

These findings suggest that it may be the *ordering of types*, rather than *rationality*, that underpins the uniqueness result. This naturally raises two questions: Could alternative, plausible heterogeneity restrictions also yield unique representations? And given that SCRUM identifies a specific RUM representing the data, what distinguishes this representation from others? What unique properties set SCRUM apart within the class of RUM representations? Our goal is to develop a novel framework to address these fundamental questions.

Our contribution is twofold. First, we show that the *ordering of types* itself is not the main driver of the uniqueness result. We demonstrate this by first establishing that there are wide-ranging models restricting heterogeneity to achieve a unique representation. These models do not require types to be ordered. To show this, we establish an intriguing connection between ordered probabilistic choice and *copula theory*. This connection provides a robust framework in which each representation of ordered probabilistic choice behavior corresponds to a copula. By leveraging established results from copula theory, we systematically examine the structure of a vast class of representations of ordered probabilistic choice behavior–a task that would otherwise be prohibitively complex.

Second, we provide functional forms to describe the "extremal" representations of an ordered probabilistic choice behavior. The resulting functional forms (copulas) act as an "identification method," that uniquely generates heterogeneous choice types and their weights. These results provide valuable tools for analysts to identify micro-level behavioral choice patterns from macro-level observable data.

We uncover a class of models that restrict heterogeneity to achieve a unique representation by connecting ordered probabilistic choice and *copula theory*. Copulas in probability theory are initially developed by Sklar (1959), who shows that any joint cumulative distribution over real numbers can be expressed as a copula composed of the marginals' distributions.[1] Thus, a copula isolates the dependence among random variables from the randomness of individual variables.

The key connection is that what we refer to as a *representation* of observed data in discrete choice is closely linked to a copula. A representation is a distribution over choice types, and must be consistent with the observed data, which is a set of probability distributions for each available choice set, such as budgets. We call it a *representation* because the *joint distribution* over choice types must reproduce the observed data as *marginal distributions*. Then, it follows from Sklar (1959) that any representation of observed data can be induced via a copula. Additionally, each copula provides a functional form that uniquely determines the heterogeneous choice types and their associated weights from given probabilistic choice dataset–a process we called an *identification method*.

[1] And uniquely so, given certain conditions on the supports of the marginals.

To explore the implications of this connection, we first observe that SCRUM or PRC representations (progressive representations) correspond to an extremal copula called *Fréchet-Hoeffding upper bound* (the min-copula), which has a particularly tractable form. This connection clarifies the uniqueness of progressive representations while providing us an explicit functional form to describe the underlying distribution of choice types and their weights, facilitating the identification.

The connection between the SCRUM and PRC representations and the *Fréchet-Hoeffding upper bound* highlights what sets the progressive representation apart from other RUM representations: For a given choice type t, the progressive representation assigns a higher total probability to types *dominating* t compared to any other representation for the same probabilistic data.[2] Consequently, the probability assigned to the choice type maximizing the underlying reference order is maximized under the progressive representation.

The uniqueness result holds as long as the probabilistic model is based on a copula. One might question if the connection between plausible probabilistic choice models and well-known copulas is merely coincidental. We suggest that many more connections remain to be discovered. For illustration, we focus on the *Fréchet-Hoeffding lower bound*. Unlike the upper bound, the lower bound is generally not a copula for more than two marginal distributions, meaning that the lower bound cannot always generate joint distributions from given marginals. Therefore, a model identified by the lower bound inherently possesses empirical content. However, like the upper bound, the corresponding representation is unique when it exists. We uncover the full empirical content of the Fréchet-Hoeffding lower bound.

To demonstrate how copula theory can uncover intriguing probabilistic choice models, we introduce the *1-mistake model* and show that it is identified by the Fréchet-Hoeffding lower bound. To motivate this model, consider a group of individuals aiming to maximize a common reference order. Occasionally, they fail to choose the optimal alternative. We term these deviations "mistakes," which may occur due to cognitive lim-

[2] A choice type *dominates* another if, in all choice sets, it consistently selects an alternative that is the same as or ranked higher than the alternative chosen by the other type according to the reference order.

itations or the use of various decision-making heuristics.[3] In an 1-mistake model, individuals are allowed to make a single mistake. This model posits that each choice type is either entirely rational (free of mistakes) or makes a mistake in a single choice set. Unlike PRC, the 1-mistake model has empirical content fully captured by a single axiom. Moreover, unlike RUM, it offers an interpretation that avoids imposing unrealistically large behavioral heterogeneity in the population. In Section 5, we present several other examples of copulas with empirical content that may be useful in applications.

Related literature

We aim to bridge the fields of ordered probabilistic choice and copula theory, offering a novel perspective on key concepts in discrete choice such as *representation* and *identification*. Ordered probabilistic choice has a long history, with roots in discrete choice (e.g, Amemiya 1981, Small 1987, Agresti 1984). Our main objective is to examine how individuals make probabilistic choices among ordered alternatives. Apesteguia et al. (2017) has reinvigorated interest in this area (Barseghyan et al. 2021, Tserenjigmid 2021, Turansick 2022, Yildiz 2022, Filiz-Ozbay & Masatlioglu 2023, Apesteguia & Ballester 2023*a*,*b*, Petri 2023, Masatlioglu & Vu 2024). This resurgence in interest stems from the growing availability of detailed choice data and theoretical models.

The literature on copulas in statistics is vast to survey here, but we emphasize that nearly every result we discuss here has a continuous counterpart in this literature. An excellent textbook treatment is provided by Nelsen (2006); Schweizer & Sklar (2005) is also a standard reference. The bounds we refer to seem to be named after the contributions by Fréchet (1935, 1951), Hoeffding (1940). Sklar's theorem appears in Sklar (1959). The results related to the one-mistake model are understood in statistics as results on negative dependence; fundamental results are due to Dall'Aglio (1972), see Lauzier, Lin & Wang (2023) for a modern treatment. Copulas are extensively utilized in econometrics and finance—Fan & Patton (2014) provide a systematic treatment. Copulas have been used in other areas of economic theory, for example, in bargaining Bastianello & LiCalzi (2019), in auction Gresik (2011), in behavioral game theory Frick, Iijima & Ishii (2022).

[3]The 1-mistake model is a special case of the models examined by Masatlioglu & Yildiz (2025).

2 Ordered Probabilistic Choice and Copulas

2.1 Preliminaries

Let X be a finite set of **alternatives**. We consider scenarios in which the alternatives possess a natural order. Examples include selecting tax policy according to the total revenue generated, choosing lotteries according to their expected monetary value, determining the number of automobiles owned, choosing the time of day for commuting, comparing insurance offers according to their deductibles, choosing public good provision, and evaluating levels of labor force participation. We call this underlying order a **reference order**, denoted by $\rhd$, which is a complete, transitive, and asymmetric binary relation over X. We write $\unrhd$ for its union with the equality relation.

Let N be the set $\{1, \ldots, n\}$, whose cardinality is denoted by n. Let $\{S_i\}_{i\in N}$ be a family of **choice sets**, where $S_i \neq \emptyset$. Notably, we do not make the full domain assumption, hence $\{S_i\}_{i\in N}$ could be a strict subset of 2^X. A (deterministic) **choice type** s is a list of alternatives $[s_1, s_2, \ldots, s_n]$ such that $s_i \in S_i$ for each $i \in N$. Let $\mathcal{S}$ and $\mathcal{S}_R$ be the set of all choice types and all *rational* types, respectively. The reference order $\rhd$ allows us to naturally compare choice types: A choice type s **dominates** another choice type s' if s_i is $\unrhd$-better than s'_i for each $i \in N$. With a slight abuse of notation, we also use the notation $\unrhd$ to describe this **dominance relation** on choice types: $s \unrhd s'$ if $s_i \unrhd s'_i$ for each $i \in N$.

Let $\rho_i(x)$ denote the probability that alternative x is chosen from the choice set S_i. Hence ρ_i is a probability distribution over S_i. A **probabilistic choice function** (`pcf`) ρ is a collection: $\{\rho_i\}_{i\in N}$. Let $\mathcal{P}$ denote the set of all `pcfs`. Given that our domain is ordered, these `pcfs` are referred to as **ordered probabilistic choices**, enabling the definition of a cumulative choice function. The **cumulative choice function** (`ccf`) associated to ρ is $P^\rho = \{P_i^\rho\}_{i\in N}$ such that

$$P_i^\rho(x) = \sum_{y\in S_i : x \unrhd y} \rho_i(y)$$

for each $x \in S_i$ and $i \in N$. We will use P and P_i instead of P^ρ and P_i^ρ when the context allows for clarity. Notably, it is the order structure that enables the unique association of a cumulative choice function with a probabilistic choice function.

As mentioned before, there is a connection between cumulative choice functions and copulas, which are flexible tools for modeling dependence among random variables. A copula creates a multivariate distribution from a given set of random variables (Nelsen 2006). Formally, a **copula** is a function $C : [0,1]^n \rightarrow \mathbb{R}$ that satisfies the following three properties. It is *grounded:* If $C(u_1, u_2, \ldots, u_n) = 0$ when any $u_i = 0$. It *has uniform margins:* If $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for any $i \in N$. Additionally, *the rectangle inequality* requires C to induce a nonnegative distribution over any n-dimensional cube, $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, ensuring that C is a valid joint distribution.[4]

Sklar's Theorem: Sklar (1959) shows that any joint cumulative distribution over the real numbers can be expressed as a copula composed with the marginal distribution functions of the joint distribution. Formally, let F be an n-dimensional cumulative distribution function (CDF) with marginal distribution functions $F_1, F_2, \ldots, F_n$. Then, there exists a copula C such that for each $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

2.2 Connection to Ordered Probabilistic Choice

To establish this connection, we first formally define when a probability distribution $\pi \in \Delta(\mathcal{S})$ over choice types represents a given pcf ρ. This requires the choice probability of x from S_i be equal to the total weights of choice types who choose x from S_i.

Definition. Let ρ be an pcf and $\pi \in \Delta(\mathcal{S})$ be a probability distribution over choice types. Then, π **represents** ρ if for each i and $x \in S_i$ we have

$$\rho_i(x) = \sum_{s : s_i = x} \pi(s).$$

Understanding of probability distributions over choice types that represent pcfs is at the heart of many exercises in probabilistic choice. This motivates us to describe a *choice*

[4]This is a variant of the inclusion-exclusion principle (or n-increasing property). The inclusion-exclusion formula ensures the C-volume of any hyperrectangle is non-negative if $\mathbf{a} \leq \mathbf{b}$. That is $\sum_{A \subseteq N} (-1)^{|A|} C(\mathbf{a}_A, \mathbf{b}_{N \setminus A}) \geq 0$ where $\mathbf{a}_A$ denotes the coordinates a_i for $i \in A$ and $\mathbf{b}_{N \setminus A}$ denotes the coordinates b_i for $i \notin A$. The alternating signs $(-1)^{|A|}$ stem from the inclusion-exclusion principle, serving to adjust for overcounting or undercounting in the joint distribution.

model as a set of $\langle \rho, \pi \rangle$ pairs where ρ is the observed data consistent with the model and π is an unobservable representation of ρ.

Definition. A (choice) **model** $\mathcal{M}$ is a set of $\langle \rho, \pi \rangle$ pairs where ρ is an `pcf` and $\pi \in \Delta(\mathcal{S})$ is a probability distribution over choice types that represents ρ.

We next describe two new objects associated with a model $\mathcal{M}$. For each `pcf` ρ, let $I_{\mathcal{M}}(\rho)$ be the set of all representations π such that $\langle \rho, \pi \rangle$ is contained in $\mathcal{M}$, i.e., $I_{\mathcal{M}}(\rho) := \{\pi | \langle \rho, \pi \rangle \in \mathcal{M}\}$. Note that if there is no representation π such that $\langle \rho, \pi \rangle \in \mathcal{M}$, then $I_{\mathcal{M}}(\rho) = \emptyset$. Let $\mathcal{P}_{\mathcal{M}}$ be the set of `pcfs` such that $I_{\mathcal{M}}(\rho) \neq \emptyset$. So, $\mathcal{P}_{\mathcal{M}}$ is the set of `pcfs` that are **consistent** with a model $\mathcal{M}$.[5]

A classical representation theorem in decision theory focuses on the properties satisfied by $\mathcal{P}_{\mathcal{M}}$. When $\mathcal{P}_{\mathcal{M}} = \mathcal{P}$, where $\mathcal{P}$ represents the set of all `pcfs`, the model has no empirical content, as it can account for every conceivable behavior. In contrast, a smaller $\mathcal{P}_{\mathcal{M}}$ enhances the model's predictive power by imposing constraints on the behaviors it can explain, thereby making it empirically meaningful.

An identification theorem examines the *size of* $I_{\mathcal{M}}(\rho)$, which is the set of representations in $\mathcal{M}$ consistent with the observed behavior ρ. A model $\mathcal{M}$ is **uniquely identified** if $|I_{\mathcal{M}}(\rho)| = 1$ for every $\rho \in \mathcal{P}_{\mathcal{M}}$. This means that for any observed behavior ρ consistent with $\mathcal{M}$, there exists a single admissible representation π that explains ρ.

We use the Random Utility Model (RUM) to illustrate the notation introduced above. RUM consists of `pcfs` that can be represented by a probability distribution over rational types. It is well known that, in general, the distribution over choice types cannot be uniquely identified from probabilistic choice data (e.g., Falmagne 1978, Fishburn 1998), i.e, $|I_{RUM}(\rho)| \neq 1$ for some $\rho \in \mathcal{P}_{RUM}$.[6]

[5]It may be tempting to refer to $\mathcal{P}_{\mathcal{M}}$ as a model. However, two distinct models, $\mathcal{M}$ and $\mathcal{M}'$, can satisfy $\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\mathcal{M}'}$ even though $\mathcal{M} \neq \mathcal{M}'$ (see footnote 6 for an example). This distinction enables a more expressive framework for differentiation.

[6]Suleymanov (2024) introduces the branch-independent RUM (BI-RUM), a model with the same explanatory power as RUM, i.e., $\mathcal{P}_{RUM} = \mathcal{P}_{BRUM}$. However, BI-RUM can be uniquely identified: $|I_{BI-RUM}(\rho)| = 1$ for each $\rho \in \mathcal{P}_{RUM}$. Additionally, $I_{BI-RUM}(\rho) \in I_{RUM}(\rho)$. Thus, while RUM and BI-RUM share the same explanatory power, they differ in their identification properties. Recent research by Caliari & Petri (2024) introduces a novel model that achieves the same explanatory power as RUM.

A key question is: What is the structure of probability distributions over choice types that represent a given pcf? Answering this question is crucial for understanding which population interpretations can be legitimately derived from observed data.

To answer this question, we rely on a *systematic* method of associating a probability distribution over choice types, a representation π, to a given pcf ρ. Determining π is generally a challenging task, often involving intricate constructions. For certain models, while the existence of such a representation is established, its precise structure remains elusive. This raises the question of whether it is possible to define a functional form–serving as an identification method–that uniquely determines π from a given pcf.

Definition. An **identification method** for a model $\mathcal{M}$ is a mapping $\mathcal{I} : \mathcal{P}_{\mathcal{M}} \to \Delta(\mathcal{S})$ such that $\mathcal{I}(\rho) \in I_{\mathcal{M}}(\rho)$ for every $\rho \in \mathcal{P}_{\mathcal{M}}$.

If $I_{\mathcal{M}}(\rho)$ contains exactly one representation for every $\rho \in \mathcal{P}_{\mathcal{M}}$, then $\mathcal{M}$ is uniquely identified. The connection between copulas and ordered probabilistic choice stems from the observation that copulas serve as concrete examples of identification methods.

We first show that a copula C gives birth to a unique probability distribution that represents ρ. Hence, each copula can be seen as an identification method. To see this, for a given copula C, we define the mapping $\mathcal{I}^C : \mathcal{P} \to \Delta(\mathcal{S})$ such that $\mathcal{I}^C(\rho)$ is the probability distribution over types induced by copula C representing ρ. Here, $\mathcal{I}^C(\rho)(s)$ is the weight assigned to the choice type $s \in \mathcal{S}$ such that the following identity holds.

$$\mathcal{I}^C(\rho)(\{s' : s \trianglerighteq s'\}) = C(P_1^{\rho}(s_1), \ldots, P_n^{\rho}(s_n)). \tag{1}$$

Identity (1) defines the probability distribution $\mathcal{I}^C(\rho)$ through its *multivariate* CDF, capturing the distribution over choice types. As in classical statistics, this representation uniquely extends to a probability measure over choice types, thereby representing the given pcf ρ. Thus, each copula has the potential to provide a functional form that uniquely determines heterogeneous choice types and their associated weights from a given probabilistic choice dataset. Therefore, each copula C induces a model $\mathcal{M}^C = \{\langle \rho, \mathcal{I}^C(\rho)\rangle | \rho \in \mathcal{P}\}$, where $\mathcal{I}^C$ is the unique identification method for $\mathcal{M}^C$.

Critically, however, the authors demonstrate that its identification framework is fundamentally distinct, revealing that the supports of RUM and their proposed representations share no common individual types.

In section 3, we illustrate that $\mathcal{I}^M$, where M stands for the min-copula, is the identification method for the progressive random choice model of Filiz-Ozbay & Masatlioglu (2023), hence $\mathcal{M}^M = \mathcal{M}_{PRC}$.

2.3 Fréchet-Hoeffding Bounds

In this subsection, we present a fundamental result in copula theory, which we later employ for our purposes. Hoeffding (1940) and Fréchet (1935, 1951) independently showed that a copula always lies between two specific bounds.

Theorem: (Fréchet-Hoeffding bounds) For each copula C,

$$\max\{\sum_{i=1}^{n} u_i + 1 - n, 0\} \leq C(u_1, u_2, \cdots, u_n) \leq \min\{u_1, u_2, \cdots, u_n\}.$$

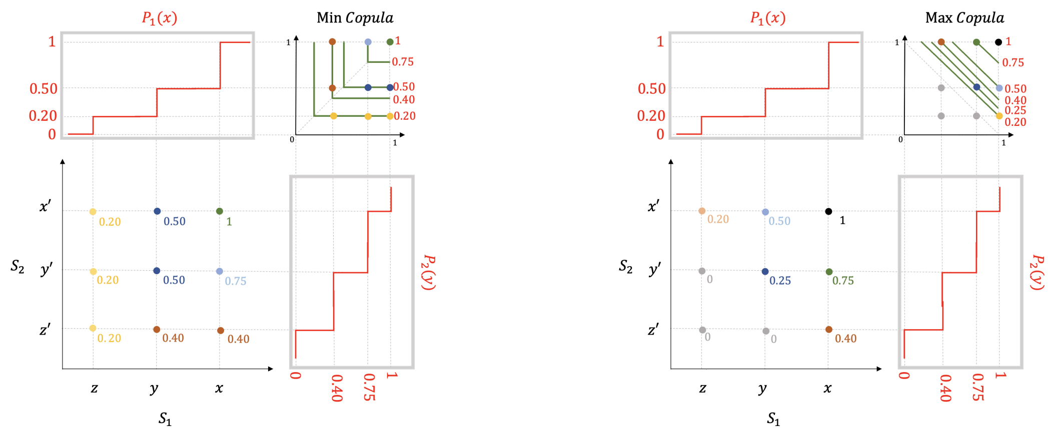
Moreover, these bounds are pointwise sharp, i.e. for each $u \in [0, 1]^n$,

$$\inf_C C(u) = \max\{\sum_{i=1}^{n} u_i + 1 - n, 0\} \text{ and } \sup_C C(u) = \min\{u_1, u_2, \cdots, u_n\}.$$

Historically, the **FH-lower bound** and the **FH-upper bound** (**min-copula**) have been denoted by W and M respectively, and we follow this notation here. While the upper bound is itself always a copula, the lower bound is only a copula in the case of $n = 2$. When $n = 2$, these two bounds correspond to distinct forms of extreme dependencies. In the case of the upper bound, the two random variables are perfectly aligned, exhibiting a property known as *comonotonicity*. Conversely, in the case of the lower bound, the two random variables move in opposite directions, exhibiting a property known as *countermonotonicity*.

Next, we illustrate how these two copulas generate two distinct representations for the same choice data. We consider two disjoint choice sets $S_1 = \{x, y, z\}$ and $S_2 = \{x', y', z'\}$ with marginal choice probabilities $\rho_1(z) = 0.20$, $\rho_1(y) = 0.30$, $\rho_2(z') = 0.40$, and $\rho_2(y') = 0.35$. We assume that the reference order is: $x \rhd y \rhd z$ and $x' \rhd y' \rhd z'$.

Figure 1 illustrates the process of combining marginal distributions for two separate choice problems into a joint distribution using the **FH-lower bound** and **FH-upper bound**. This approach ensures that the resulting probability structure aligns with the



(a) *M: The FH-upper bound*

(b) *W: The FH-lower bound*

Figure 1: Construction of cumulative distributions over deterministic choice functions for Fréchet-Hoeffding bounds involving two choice sets: $S_1 = \{x, y, z\}$ and $S_2 = \{x', y', z'\}$. The marginal choice probabilities are $\rho_1(z) = 0.20$, $\rho_1(y) = 0.30$, $\rho_2(z') = 0.40$, and $\rho_2(y') = 0.35$. The left panel presents the unique cumulative distribution for M. The cumulative probabilities of types below $[z, z']$ and $[y, y']$ are 0.20 and 0.50, respectively. The right panel depicts the unique distribution for W. The cumulative probabilities for types below $[z, z']$ and $[y, y']$ are now 0 and 0.25.

given marginal probabilities and the reference order. Each figure contains two rectangles representing the cumulative marginal distributions for each choice set. The corresponding copula is then applied to determine the probability levels, which are displayed in the top right corner of each figure. These values are subsequently mapped onto the distribution over choice types, as shown in the bottom left corner. The numerical values at each point represent the resulting joint probabilities.

We first calculate the cumulative marginal distributions. We have $P_1(z) = 0.20$, $P_1(y) = 0.50$, and $P_1(x) = 1.00$. Similarly, $P_2(z') = 0.40$, $P_2(y') = 0.75$, and $P_2(x') = 1.00$. Note that P_1 and P_2 are the same cumulative marginal distributions in both panels since they are based on the same choice data. In the left panel, we calculate the corresponding cumulative joint distribution using M (the FH-upper bound). For example, $M(P_1(y), P_2(z')) = M(0.50, 0.40) = 0.40$ and $M(P_1(z), P_2(x')) = M(0.20, 1.00) = 0.20$. Hence, the cumulative probabilities of the types below $[y, z']$ and $[z, x']$ are 0.40 and 0.20, respectively. On the other hand, the corresponding cumulative joint distribution using W (the FH-upper bound) are $W(P_1(y), P_2(z')) = W(0.50, 0.40) = 0$ and

$W(P_1(z), P_2(x')) = W(0.20, 1.00) = 0.20$, displayed on the right panel. Then, the cumulative probabilities of the types below $[y, z']$ are zero. This implies that the probability of the type $[z, z']$ is also zero.

Figure 2 provides the unique weights associated with each type according to W and M. This is based on the cumulative distributions provided in Figure 1. Since $M(P_1(z), P_2(z')) = M(P_1(z), P_2(x')) = 0.20$, the representation of M assigns a probability of 0.20 for the type $[z, z']$, while $[z, y']$ and $[z, x']$ have probability of 0, which is illustrated on the left panel. Given that the cumulative probability for $[y, z']$ is 0.40, we assign a probability of 0.20 to $[y, z']$ by subtracting the weight of $[z, z']$.

	M: The FH-upper bound				
Types	$[z, z']$	$[y, z']$	$[y, y']$	$[x, y']$	$[x, x']$
Weights	0.20	0.20	0.10	0.25	0.25
	W: The FH-lower bound				
Types	$[z, x']$	$[y, x']$	$[y, y']$	$[x, y']$	$[x, z']$
Weights	0.20	0.05	0.25	0.10	0.40

Table 1: Representations based on the FH-upper bound and the FH-lower bound.

Table 1 presents these two representations. Importantly, the types in the support of M are arranged in a monotonic order, forming a path from $[z, z']$ to $[x, x']$. In contrast, most of the types in the support of W are not comparable. Additionally, while M assigns a weight of 0.25 to the highest type $[x, x']$, W assigns it a weight of zero.

3 Progressive random choice and the FH-upper bound

In this section, we first introduce the model of Filiz-Ozbay & Masatlioglu (2023) (FM) and show that the min-copula M is the identification method for this model. In addition, we show that the connection to copula theory reveals an interesting aspect of their model that was unknown.

FM introduces an ordered probabilistic choice model in which types are ranked based on a fixed characteristic. For instance, consider a set of policies differing in their levels of environmental friendliness. Types are indexed according to their degree of environ-

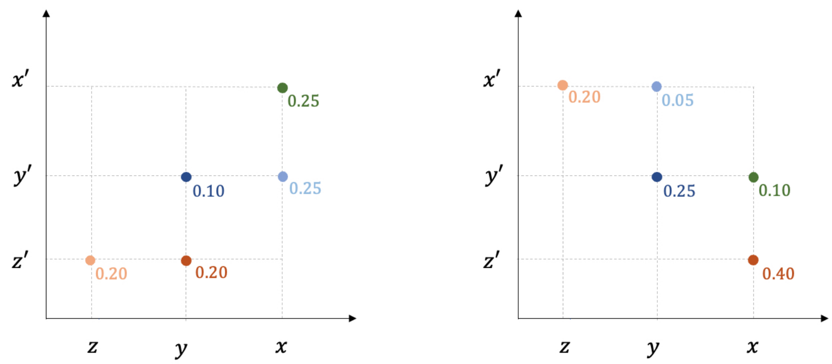


Figure 2: The representation of the distribution over deterministic choice functions for Fréchet-Hoeffding bounds shows two panels. The left panel is based on the upper bound, illustrating the distribution using the min-copula. The right panel depicts the weights induced by the lower bound.

mental caution. Under this model, a type with a higher index will not choose a less environmentally friendly policy than the one chosen by a lower-indexed type when faced with the same choice problem. This model is called the Progressive Random Choice.

Formally, a set of distinct choice types $\{s^1, \dots, s^T\}$ is **progressive** with respect to $\rhd$ if $s_i^t \unrhd s_i^{t+1}$ for each $i \in N$ and $t \in \{1, \dots, T-1\}$. The progressive structure reduces the heterogeneity of types into a single dimension since choice types gradually become more and more aligned with the choice induced by $\rhd$. For a given reference order $\rhd$, an `pcf` ρ is a **progressive random choice** (PRC) if there exists a probability distribution π over $\rhd$-progressive deterministic choice types such that π represents ρ. Formally,

$$PRC := \{\langle \rho, \pi \rangle \mid \pi \text{ has a progressive support and represents } \rho\}.$$

In their main result, FM shows that every probabilistic choice has a unique PRC representation denoted by π^{ρ}_{PRC}, i.e., $\mathcal{P}_{PRC} = \mathcal{P}$ and $I_{PRC}(\rho) = \{\pi^{\rho}_{PRC}\}$.

We now illustrate that their results follow from the existence of the min-copula. First, note that the min-copula only assigns positive weights to a set of deterministic choice types that are comonotonic, and thus progressive. Furthermore, this representation is always unique. Therefore, the min-copula is an identification method for PRC. Since the min-copula remains a copula regardless of n, $\mathcal{I}^M(\rho)$ is a probability distribution for any `pcf` ρ. Hence, $\mathcal{P}_{PRC} = \mathcal{P}$. This discussion establishes Theorem 1 of FM, highlighting the significant connection between the progressive structure and the min-copula. Additionally, the min-copula provides an explicit functional form for calculating the weights assigned to each deterministic choice function.

The connection between copula theory and ordered probabilistic choice further reveals another unknown aspect of the PRC representation: For a given choice type s, the PRC representation π^{ρ}_{PRC} assigns a higher probability to choice types dominating s compared to any other probability distribution π over choice types that represents ρ. It is also true that the PRC representation of ρ assigns a higher probability to the choice types weakly dominated by s compared to any representation π of ρ. We next formally state and prove this result.

Proposition 1 *Let ρ be an `pcf`, and let π be a probability distribution over choice types that represents ρ. Then, for each choice type $s \in \mathcal{S}$, the PRC representation of ρ, denoted by π^{ρ}_{PRC}, satisfies the inequalities*

$$\pi^{\rho}_{PRC}(\{s' : s' \trianglerighteq s\}) \geq \pi(\{s' : s' \trianglerighteq s\}) \text{ and } \pi^{\rho}_{PRC}(\{s' : s \trianglerighteq s'\}) \geq \pi(\{s' : s \trianglerighteq s'\}).$$

Proof: For the second inequality, since the PRC representation π^{ρ}_{PRC} corresponds to the min-copula, it is sufficient to show that $\min\{P^{\rho}_1(s_1), \ldots, P^{\rho}_n(s_n)\} \geq \pi(\{s' : s \trianglerighteq s'\})$. To see this, first note that by Sklar's Theorem there exists a copula C such that $\pi(\{s' : s' \trianglerighteq s\}) = C(P^{\rho}_1(s_1), \ldots, P^{\rho}_n(s_n))$. Then, since min-copula is the FH-upper bound, we have $\min\{P^{\rho}_1(s_1), \ldots, P^{\rho}_n(s_n)\} \geq C(P^{\rho}_1(s_1), \ldots, P^{\rho}_n(s_n))$.

To show that the first inequality holds, consider $\trianglerighteq^{-1}$ which is the inverse of the reference relation $\trianglerighteq$. Let $s, s' \in \mathcal{S}$ be any two choice types that are assigned positive probability by $\mathcal{I}^M(\rho)$ under the reference relation $\trianglerighteq^{-1}$. Since s and s' are comparable under $\trianglerighteq^{-1}$, they are also comparable under $\trianglerighteq$. Then, since the PRC representation is unique, we must have $\pi^{\rho}_{PRC} = \mathcal{I}^M(\rho)$ (this equivalence with respect to $\trianglerighteq$ and $\trianglerighteq^{-1}$ that holds for the min-copula may not hold for other copulas). Therefore, applying the second inequality under $\trianglerighteq^{-1}$ yields $\pi^{\rho}_{PRC}\left(\{s' : s \trianglerighteq^{-1} s'\}\right) \geq \pi\left(\{s' : s \trianglerighteq^{-1} s'\}\right)$, which establishes the second inequality.

An immediate implication of Proposition 1 is that the probability assigned to the choice type maximizing the underlying reference order is maximized by π^{ρ}_{PRC}. Similarly, π^{ρ}_{PRC} maximizes the probability assigned to the choice type minimizing the underlying reference order. To gain intuition, recall that for each s, s' in the support of π^{ρ}_{PRC}, we have $s \trianglerighteq s'$ or $s' \trianglerighteq s$. Proposition 1 sharpens this intuition by requiring that, for *every* s, the probability of obtaining an s' that is comparable to s (in either direction) under $\trianglerighteq$ is maximized.

As noted in the introduction, SCRUM identifies a specific RUM representation as its unique form. Next, we formally define SCRUM.

$$SCRUM := \{\langle \rho, \pi \rangle \mid \pi \in \Delta(\mathcal{S}_R) \text{ has a progressive support and represents}\}.$$

The distinction between PRC and SCRUM lies in the support of their representations. While PRC can accommodate any probabilistic choice, SCRUM imposes an additional restriction where each type must be rational, providing empirical content for SCRUM. This restriction gives SCRUM its predictive power: $\mathcal{P}_{SCRUM} \subset \mathcal{P}_{RUM} \subset \mathcal{P}_{PRC}$. Indeed, Apesteguia et al. (2017) show that choice data satisfies both *centrality* and *regularity* if and only if it has a SCRUM representation.[7] This implies that the min-copula assigns positive weights only to the rational types if and only if *centrality* and *regularity* are satisfied. On the other hand, for each $\rho \in \mathcal{P}_{SCRUM}$, $I_M(\rho) = I_{PRC}(\rho) = I_{SCRUM}(\rho)$. Thus, the min-copula serves as an identification method for SCRUM.

With the above observation in hand, we can now examine what sets the SCRUM representation apart from other RUM representations. Proposition 1 provides answers to these questions. SCRUM selects a representation from the set of RUM representations in which the probability assigned to the choice type that maximizes the underlying reference order is maximized.

The min-copula acts as an identification method for any model that incorporates a progressive structure along with certain additional support conditions. In this vein, FM introduces two special cases of PRC: (i) "less-is-more" and (ii) "no-simple-mistakes." These special cases, like SCRUM, introduce additional restrictions on choice types. The less-is-more model allows only types that make fewer mistakes with smaller sets. The no-simple-mistakes model ensures that each type does not choose an option in ternary

[7] To state their axioms, we use their notation where $\rho(x, A)$ represents the choice probability of x in A. Then, *Regularity*: If $x \in B \subset A$, then $\rho(x, A) \leq \rho(x, B)$. *Centrality*: If $x \triangleright y \triangleright z$ and $\rho(y, \{x, y, z\}) > 0$, then $\rho(x, \{x, y\}) = \rho(x, \{x, y, z\})$ and $\rho(z, \{z, y\}) = \rho(z, \{x, y, z\})$.

comparisons that has never been chosen in binary comparisons. FM provides behavioral postulates that characterize them. The min-copula can be used to identify the weights in all these models.

Next, we report another class of models that can be identified using the min-copula. Yildiz (2022) defines a set of choice types $\mathcal{S}^*$ as *self-progressive* if every probabilistic choice function that can be represented as a probability distribution over choice types in $\mathcal{S}^*$ can also be progressively represented using (possibly different) choice types within $\mathcal{S}^*$. Yildiz (2022) characterizes the comprehensive family of self-progressive choice models by establishing that $\mathcal{S}^*$ is self-progressive if and only if $\langle \mathcal{S}^*, \trianglerighteq \rangle$ is a *lattice*: For every pair of choice types in $\mathcal{S}^*$, the choice types corresponding to their *join*—formed by collecting the $\trianglerighteq$-better choices—and their *meet*—formed by collecting the $\trianglerighteq$-worse choices—are also contained in $\mathcal{S}^*$. It follows directly that, for any set of types $\mathcal{S}^*$ such that $\langle \mathcal{S}^*, \trianglerighteq \rangle$ is a lattice, the min-copula serves as an identification method for the associated model $\mathcal{M}_{\mathcal{S}^*} := \{\langle \rho, \pi \rangle \mid \pi \in \Delta(\mathcal{S}^*) \text{ that represents } \rho\}$.

4 Probabilistic choice induced by the FH-lower bound

We have shown that the min-copula uniquely identifies all special cases of PRC. At first glance, the connection between probabilistic choice and a well-known copula might appear unexpectedly coincidental. However, we now contend that copula theory can serve as a powerful tool for uncovering plausible choice models. To illustrate this, we first introduce a choice model identified by the FH-lower bound. Then, we explore the full empirical implications of the FH-lower bound.

We first remind you that the FH-lower bound W is generally not a copula for $n > 2$. This means that naively applying W to an `pcf` may not result in a probability distribution over choice types. To cover such cases, we leverage the concept of *quasi-copula* formulated by Alsina, Nelsen & Schweizer (1993) in the bivariate case, and Nelsen, Quesada-Molina, Schweizer & Sempi (1996) for the general case. Quasi-copulas generalizes copulas, relaxing some of their strict requirements while still preserving key properties that make them useful in modeling dependence structures. Namely, the rectangular inequality requirement of a copula is replaced by a Lipschitz condition.[8] Now, we define when a quasi-copula identifies an `pcf`.

Definition. A **quasi-copula** Q **identifies** an `pcf` ρ if $\mathcal{I}_Q^\rho$ is a probability distribution over $\Delta(\mathcal{S})$ and represents ρ such that for each choice type s,

$$\mathcal{I}^Q(\rho)(\{s' : s \trianglerighteq s'\}) = Q(P_1^\rho(s_1), \ldots, P_n^\rho(s_n)). \tag{2}$$

Identity (2) holds for any `pcf` ρ. However, $\mathcal{I}^Q(\rho)$ might not be a probability distribution. Mirroring the properties of copulas, the corresponding representation is unique whenever it exists. A model induced by a quasi-copula Q is well-defined provided that $\mathcal{I}^Q(\rho)$ forms a valid probability distribution. The set $\mathcal{M}^Q$ consists of all pairs $\langle \rho, \mathcal{I}^Q(\rho)\rangle$ for which $\mathcal{I}^Q(\rho)$ is a probability distribution.

$$\mathcal{M}^Q := \{\langle \rho, \mathcal{I}^Q(\rho)\rangle \mid \mathcal{I}^Q(\rho) \text{ is a probability distribution that represents } \rho\}.$$

Let $\mathcal{P}_{\mathcal{M}^Q}$ denote the set of all `pcfs` identified by Q. Notably, if Q is a copula, then $\mathcal{P}_{\mathcal{M}^Q} = \mathcal{P}$. We interpret any `pcf` for which the application of Q fails to yield a probability distribution as being *ruled out* by $\mathcal{M}^Q$ (or, equivalently, by Q itself). Finally, observe that, by construction, $\mathcal{I}^Q$ uniquely determines $\mathcal{M}^Q$.

4.1 A model identified by the FH-lower bound

As illustrated in the last section, the model induced by M corresponds to PRC. We argue that copula theory can uncover plausible and intriguing probabilistic choice models. We first introduce the 1-mistake model, and show that it is identified by the FH-lower bound. We then investigate the behavioral content of $\mathcal{M}_W$.

Consider a set of individuals aiming to maximize their reference order $\triangleright$. However, at times, they may deviate from selecting the best available alternative. We refer to these deviations as “mistakes” that reflect cognitive limitations or the use of different decision heuristics by individuals. In our 1-mistake model, individuals are allowed to make a single mistake, getting the choice incorrect for at most one choice set. This model posits that each type is either entirely rational (free of mistakes) or makes a mistake in a

[8] A function Q satisfies the *Lipschitz condition* if $|Q(\mathbf{u}) - Q(\mathbf{u}')| \leq \sum_i |u_i - u_i'|$ for every $\mathbf{u}, \mathbf{u}' \in [0,1]^n$.

single-choice set. Formally, a choice type s is **near $\triangleright$-optimal** if there exists at most one choice set S_i such that s_i differs from the $\triangleright$-best element in S_i. Let $\mathcal{N}_{\triangleright}$ denote all **near $\triangleright$-optimal** choice types.

To provide a simple example, let $S = S_1 \times S_2 \times S_3$ where $S_1 = \{x, y, z\}$, $S_2 = \{x, y\}$, and $S_3 = \{x, z\}$. We assume that the reference order is $x \triangleright y \triangleright z$. Then, for example, $[z, y, x]$ represents the choice function choosing z, y, and x from S_1, S_2, and S_3. Each of $[x, x, x]$, $[y, x, x]$, $[z, x, x]$, $[x, x, z]$, and $[x, y, x]$ are nearly $\triangleright$-optimal choice types. Now, we can define the 1-mistake model. An `pcf` ρ has a 1**-mistake** representation with respect to $\triangleright$ if there exits a probability distribution π over near $\triangleright$-optimal choice types that generates ρ, i.e., 1**-mistake** $:= \{\langle \rho, \pi \rangle \mid \pi \in \Delta(\mathcal{N}_{\triangleright})$ that represents $\rho\}$.

Unlike PRC, 1-mistake model possesses empirical content, i.e., $\mathcal{P}_{\textbf{1-mistake}} \neq \mathcal{P}$. In the first part of the following result, we present a postulate that encapsulates the behaviors induced by this model. This postulate asserts that the total probability of mistakes must be less than 1. The second part of the result establishes that the 1-mistake model is identified by the FH-lower bound.

Proposition 2 *Let ρ be an `pcf` and $\bar{s}_i$ be the $\triangleright$-best alternative in S_i. Then,*

i. $\rho \in \mathcal{P}_{\textbf{1-mistake}}$ if and only if $\sum_{i \in N}(1 - \rho_i(\bar{s}_i)) \leq 1$.

ii. If $\rho \in \mathcal{P}_{\textbf{1-mistake}}$, then ρ is identified by the FH-lower bound.

Proof: i. If ρ is a 1-mistake model then it immediately follows that $\sum_{i \in N}(1 - \rho_i(\bar{s}_i)) \leq 1$. Conversely, let ρ be an `RCF` such that $\sum_{i \in N}(1 - \rho_i(\bar{s}_i)) \leq 1$. Then, for each $i \in N$ and $s_i \neq \bar{s}_i$, define the choice type s^i such that $s^i_i = s_i$ and $s^i_j = \bar{s}_j$ for each $j \neq i$. Let $\bar{s}$ be the choice type such that $s_i = \bar{s}_i$ and for each $i \in N$. Now, define a distribution π over choice types such that $\pi(s^i) = \rho_i(s_i)$ and $\pi(\bar{s}) = 1 - \sum_{i \in N}(1 - \rho_i(\bar{s}_i))$, which is nonnegative by our assumption. Thus, π generates ρ, and has a support consisting of near $\triangleright$-optimal choice types.

ii. Let ρ be a 1-mistake model, and let s be a choice type such that s_i is the alternative chosen from S_i for each $i \in N$. Suppose that $s = \bar{s}$. Then, $\max\{\sum_{i=1}^n P_i^\rho(\bar{s}_i) + 1 - n, 0\} = \max\{1, 0\} = 1$. Suppose that there exists unique $i \in N$ such that $s_i < \bar{s}_i$. Then,

$\max\{P_i^\rho(s_i) + \sum_{j\neq i} P_j^\rho(\bar{s}_j) + 1 - n, 0\} = \max\{P_i^\rho(s_i), 0\} = P_i^\rho(s_i)$. Finally, suppose that there exist at least two $i, j \in N$ with $s_i < \bar{s}_i$ and $s_j < \bar{s}_j$. Let $\bar{s}_i - 1$ be the element that is immediately $\triangleright$-worse than s_i. Recall that by Part i, $\sum_{i=1}^n (1 - \rho_i(\bar{s}_i)) \leq 1$. Then, we have $\sum_i P_i^\rho(\bar{s}_i - 1) \leq 1$. It follows that $\sum_{i=1}^n P_i^\rho(s_i) \leq \sum_{i=1}^n P_i^\rho(\bar{s}_i - 1) + (n-2) \leq 1 + n - 2$, as there are at most $(n-2)$ components with $s_i = \bar{s}_i$, each of which put at most probability 1 on $\bar{s}_i$. Thus, $\sum_i P_i^\rho(s_i) + 1 - n \leq 0$ and $\max\{\sum_i P_i^\rho(s_i) + 1 - n, 0\} = 0$.

4.2 The Full Empirical Content of the FH-lower Bound

An intriguing question is whether the FH-lower bound can identify additional models beyond the 1-mistake model. To address this, we examine the full empirical implications of the FH-lower bound and characterize the class of choice models it encompasses. This analysis allows us to establish a counterpart to the equivalence between progressive random choice and the FH-upper bound for the FH-lower bound. The notion of being 1-mistake away from a given choice type is critical for our result.

Definition. A choice type s is **1-mistake away from** s^* if there exists at most one $i \in N$ such that $s_i \neq s_i^*$.

In the 1-mistake model, each admissible choice type is one mistake away from the rational type that maximizes the reference relation. In contrast, our next result shows that a model identified by the FH-lower bound permit choice types that are one mistake away from two specific choice types. Moreover, any `pcf` identified by the FH-lower bound belongs to this class, provided that the `pcf` selects at least two alternatives in at least three choice sets with positive probability.

We need to establish a few notations to present our result. Let ρ be an `RCF`. Then, for each $i \in N$, let $S_i^+ = \{x \in S_i : \rho_i(x) > 0\}$ and $\bar{s}_i^\rho$ ($\underline{s}_i^\rho$) be the $\triangleright$-best(worst) alternative in S_i^+. Let $\bar{s}^\rho = [\bar{s}_1^\rho, \ldots, \bar{s}_n^\rho]$ and $\underline{s}^\rho = [\underline{s}_1^\rho, \ldots, \underline{s}_n^\rho]$.

Proposition 3 *An* `pcf` ρ *is identified by the FH-lower bound* W *if and only if*

I. *there exists a probability distribution over choice types that are 1-mistake away from either* $\bar{s}^p$ *or* $\underline{s}^p$*, or*

II. *there exist* $i, j \in N$ *such that if* $k \in N \setminus \{i, j\}$*, then* $\rho_k(y) = 1$ *for some* $y \in S_k$*.*

Proof: Please see Section 7.

To interpret this result, consider a population of agents whose choices are centered around a salient choice type s^*, meaning that each type differs from s^* in at most one choice set. This indicates a relatively homogeneous population. Our characterization reveals that, under mild conditions (specifically when condition II fails to hold), being identifiable by the FH-lower bound necessitates that the salient choice type s^* be an extreme choice type according to the reference order.

An earlier study Dall'Aglio (1972) establishes the counterpart of our Proposition 3 for continuous random variables. In Section 7, we provide a distinct self-contained proof.

5 Other examples

The set of possible copulas is very rich (Nelsen 2006). Hence it is beyond the scope of this paper to list them. Instead, we provide several interesting (quasi-)copulas, which could be useful for generating and identifying new models of ordered probabilistic choice. We will use the FH-lower bound (W) and FH-upper bound (M) in our constructions.

Example 1 (Independent Copula) The independent copula, denoted by Π, represents a structure where each individual variable is independent of the others. It is arguably the simplest copula, defined as $\Pi(u_1, u_2, \cdots, u_n) := \Pi_i \; u_i$. The independent copula remains unaffected by the reference order and finds applications in diverse fields, including statistical modeling, finance, and machine learning.

Example 2 (Fréchet Copula Family) The family $\{C_\alpha\}$ is a one-parameter collection of quasi-copulas, representing a version of the Fréchet copula family Fréchet (1958). Each C_α is a linear combination of the Fréchet-Hoeffding (FH) lower and upper bounds, satisfying $C_0 = W$ and $C_1 = M$:

$$C_\alpha(u_1, u_2, \ldots, u_n) := \alpha M(u_1, u_2, \ldots, u_n) + (1-\alpha) W(u_1, u_2, \ldots, u_n).$$

In general, C_α is not a copula. However, similar to the FH lower bound W, one can derive conditions on the marginal distributions under which C_α induces a valid joint distribution. As established in Proposition 2, $\mathcal{I}^W$ yields non-negative weights. Given that C_α is a convex combination of W and M, the condition required for $\mathcal{I}^{C_\alpha}$ is necessarily weaker than that for $\mathcal{I}^W$ and varies monotonically with α. Finally, the support of $\mathcal{I}^{C_\alpha}$ corresponds to the union of progressive and near-optimal choice types.

Example 3 (Threshold Copula Family) This family is inspired by an example in Nelsen (2006). The threshold copula C_t is constructed as a mixture of the Fréchet-Hoeffding lower bound W and upper bound M, determined by a threshold level t. Specifically, C_t behaves like M when probabilities are below the threshold, and like W when all variables exceed the threshold. Formally,

$$C_t(u_1, u_2, \ldots, u_n) := \begin{cases} \max\left\{\sum_{i=1}^n u_i + 1 - n,\ t\right\} & \text{if } \forall i \in N, u_i \geq t, \\ M(u_1, u_2, \ldots, u_n) & \text{otherwise.} \end{cases}$$

Similar to Example 2, C_t is a combination of W and M, satisfying $C_0 = W$ and $C_1 = M$. However, C_t is not a copula in general, and the support of $\mathcal{I}^{C_t}$ may extend beyond the union of progressive and near-optimal choice types.

New copulas can be constructed by combining existing ones. A natural way to build multidimensional copulas is to use two dimensional copulas as building blocks, coupling them to form higher-dimensional structures. The following examples illustrate this idea.

Example 4 (Nested Aggregation) New quasi-copulas can be constructed by grouping choice problems into clusters and applying distinct aggregation functions to model within-group and between-group interactions. For instance, one can define:

$$C(u_1, u_2, \ldots, u_n) := M\big(W\big(W(u_1, u_2), u_3\big), \ldots, u_n\big).$$

This nested structure combines W and M to capture varying dependence patterns across different choice sets.

Example 5 (Difficulty-Based Grouping) Consider a setting with three groups of choice sets, categorized by difficulty: $G_1 = \{1, \ldots, k_1\}$ (easy), $G_2 = \{k_1 + 1, \ldots, k_2\}$ (medium), and $G_3 = \{k_2 + 1, \ldots, n\}$ (hard). A natural way to aggregate these groups is to apply the function M within each group and then W across groups:

$$C(u_1, u_2, \ldots, u_n) := W\big(M(u_1, \ldots, u_{k_1}),\ M(u_{k_1+1}, \ldots, u_{k_2}),\ M(u_{k_2+1}, \ldots, u_n)\big).$$

Note, however, that W may not necessarily be a copula in this construction.

6 Conclusion

This paper establishes a novel connection between ordered probabilistic choice and copula theory, providing a new framework for representation and identification in discrete choice models. By leveraging copulas, we show that heterogeneous choice types and their associated weights can be uniquely identified from observed data, offering analysts explicit functional forms for interpretation. Our analysis of the Fréchet-Hoeffding bounds reveals the distinctive role of the min-copula in progressive random choice models and the empirical content of the FH-lower bound in capturing bounded rationality, as illustrated by the 1-mistake model. Overall, the copula-based approach opens new avenues for constructing and analyzing rich classes of probabilistic choice models explaining behavioral heterogeneity.

We believe this paper merely scratches the surface of the vast potential offered by copula theory. We strongly encourage other researchers to explore and expand the frontiers of this promising research area.

7 Proof of Proposition 3

For each $s_i \in S_i$, we denote the element that is immediately $\rhd$-worse (better) than s_i by $s_i - 1$ ($s_i + 1$) whenever it exists (e.g., for $x \rhd a \rhd b \rhd c \rhd y$, if $s_i = b$, then $s_i + 1 = a, s_i - 1 = c$). We denote the set of all choice types by S, where $S = \prod_{i \in N} S_i$, and $S_{-j} = \prod_{i \in N \setminus \{j\}} S_i$ for each $j \in N$. Let $\bar{s}$ ($\underline{s}$) be the choice type such that $s_i = \bar{s}_i$ ($s_i = \underline{s}_i$) for each $i \in N$. For each $s, s' \in S$ and $M \subset N$, let sMs' be the element of S that copies s for the components in M, and s' for the components in $N \setminus M$.

Let ρ be an `pcf` that is identified by the W. The next lemma establishes that there is no strictly dominated choice type in the support of π^ρ_W, denoted by π^ρ_{W+}. That is, there exist no two choice types $s, s' \in \pi^\rho_{W+}$ such that $s_i \rhd s'_i$ for each $i \in N$.

Lemma 7.1 *Let ρ be an RCF that is identified by W. Then, the set of choice functions that appear in the support of π^ρ_W is an $\rhd$-antichain.*

Proof: By contradiction, suppose that there exist $s, s' \in \pi^\rho_{W+}$ such that $s_i \rhd s'_i$ for each $i \in N$. Let F^ρ_W be the CDF associated with W and ρ. That is, $F^\rho_W(s) = W(P^\rho_1(s_1), \dots, P^\rho_n(s_n))$

for each $s \in S$ (equivalently $F^\rho_W(s) = \sum_{t: s \trianglerighteq t} \pi^\rho_W(t)$). Now, let $[s-1]$ be the element of S such that $[s-1]_i = s_i - 1$ for each $i \in N$.[9] Then,

$$\pi^\rho_W(s) = \sum_{M \subseteq N} (-1)^{|M|}\, F^\rho_W([s-1]Ms). \tag{3}$$

Since s' is in the support of π^ρ_W, for each $t \in S$ such that $s \trianglerighteq t \trianglerighteq s'$, we have $F_W(t) = \sum_i P^\rho_i(t_i) - n + 1 > 0$. Note that $s \trianglerighteq [s-1]Ms \trianglerighteq s'$ for every $M \subseteq N$. It follows that

$$\pi^\rho_W(s) = \sum_{M \subseteq N} (-1)^{|M|} \left(\sum_{i \in M} P^\rho_i(s_i) + \sum_{i \in N \setminus M} P^\rho_i(s_i - 1) - n + 1 \right) \tag{4}$$

which can be decomposed into two components each of which equals zero by the principle of inclusion and exclusion.

The following result improves Lemma 7.1 by stating that any two choice functions in π^ρ_{W+} can differ at most in two choice sets. Moreover, if they deviate at two choice sets, say S_i and S_j, then we must have $s_i \triangleright s'_i$ and $s'_j \triangleright s_j$.

Lemma 7.2 *If s and s' are in the support of π^ρ_W, then*

i. $|\{i : s_i \neq s'_i\}| \leq 2$*, and*

ii. if $|\{i : s_i \neq s'_i\}| = 2$*, then* $s_j \triangleright s'_j$ *and* $s'_k \triangleright s_k$ *for some* $j, k \in N$.

Proof: For clarity, we replace π^ρ_W with π and F^ρ_W with F. Then, since π generates ρ, we can rewrite $F(s) = 1 - n + \sum_{i \in N} P^\rho_i(s_i)$ as

$$F(s) = 1 - n + \sum_{i \in N} \sum_{s_i \trianglerighteq y_i} \sum_{z_{-i} \in S_{-i}} \pi(y_i, z_{-i}). \tag{5}$$

Next, for each $k \in \{1, \ldots, n\}$, let $S^k = \{t \in S : k = |\{i : s_i \trianglerighteq t_i\}|\}$. Since, by Lemma 7.1, there is no element $t^* \in \pi^\rho_{W+}$ with $s \triangleright t^*$, we get

$$F(s) = 1 - n + n \sum_{t \in S^n} \pi(t) + \sum_{k=1}^{n-1} k \sum_{t \in S^k} \pi(t). \tag{6}$$

[9] We will use the notation $[s-1]Ms$ for the element of S for which when $i \in M$, $([s-1]Ms)_i = [s-1]$, and otherwise for $i \notin M$, $([s-1]Ms)_i = s$.

Since, by definition, $F(s) = \sum_{t \in S^n} \pi(t)$, it follows that

$$n - 1 = (n-1) \sum_{t \in S^n \cup S^{n-1}} \pi(t) + \sum_{k=1}^{n-2} k \sum_{t \in S^k} \pi(t). \tag{7}$$

Now, if $\pi(t) > 0$ for some $t \in S^k$ where $k < n-1$, then this equality fails to hold. Therefore, $\pi(t) > 0$ only if $|\{i : s_i \trianglerighteq t_i\}| \geq n-1$. It follows that $|\{i : s_i \trianglerighteq s'_i\}| \geq n-1$. Symmetrically, $|\{i : s'_i \trianglerighteq s_i\}| \geq n-1$. Thus, we conclude that i. and ii. hold.

Our next lemma provides the last stepping stone to prove Proposition 3.

Lemma 7.3 *Let ρ be an RCF that is identified by W. Suppose that ρ fails to satisfy part II of Proposition 3. Let $\alpha, \beta \in \pi^{\rho}_{W+}$ such that $\alpha_i \triangleright \beta_i$ and $\beta_j \triangleright \alpha_j$. Then, there exists $\gamma \in \pi^{\rho}_{W+}$ and $k^* \in N$ such that*

1. *$\gamma_{k^*} \neq \alpha_{k^*} = \beta_{k^*}$, and*
2. *If $\alpha_{k^*} \triangleright \gamma_{k^*}$ then α and β are 1-mistake away from $\bar{s}^p$; if $\alpha_{k^*} \triangleleft \gamma_{k^*}$ then α and β are 1-mistake away from $\underline{s}^p$.*

Proof: Note that, by Lemma 7.2 part i, we have $\alpha_k = \beta_k$ for each $k \in N \setminus \{i, j\}$. Since II fails to hold, there exists $k^* \in N \setminus \{i, j\}$ such that $p_{k^*}(\alpha_{k^*}) < 1$. It follows that there exists $\gamma \in \pi^{\rho}_{W+}$ such that $\gamma_{k^*} \neq \alpha_{k^*} = \beta_{k^*}$. Thus, 1 holds. Next, suppose w.l.o.g. that $\alpha_{k^*} \triangleright \gamma_{k^*}$. Then, by Lemma 7.2 part ii, for each $s \in \{\beta, \alpha\}$, it must be that $\gamma_i \trianglerighteq s_i$ and $\gamma_j \trianglerighteq s_j$. Therefore, $\gamma_i \trianglerighteq \alpha_i \triangleright \beta_i$ and $\gamma_j \trianglerighteq \beta_j \triangleright \alpha_j$. Then, by Lemma 7.2 part i, for γ not to differ from α and β on more than two components, we must have $\gamma_i = \alpha_i$, $\gamma_j = \beta_j$, and $\beta_k = \alpha_k = \gamma_k$ for each $k \in N \setminus \{i, j, k^*\}$.

In what follows, we show that α and β are 1-mistake away from $\bar{s}^p$ (had we supposed that $\gamma_{k^*} \triangleleft \beta_{k^*} = \alpha_{k^*}$, we would be showing that α and β are 1-mistake away from $\underline{s}^p$). That is, we claim that $\alpha_i = \bar{s}^{\rho}_i$, $\beta_j = \bar{s}^{\rho}_j$, and $\bar{s}^{\rho}_k = \alpha_k = \beta_k$ for each $k \in N \setminus \{i, j\}$. By contradiction, suppose that this fails to hold for some $l \in N$.

Case 1: Suppose that $l \in N \setminus \{i, j\}$. It follows that there exists $\delta \in \pi^{\rho}_{W+}$ such that $\delta_l \triangleright \beta_l = \alpha_l$. By Lemma 7.2 part ii, for each $s \in \{\alpha, \beta\}$, it must be that $\delta_i \trianglelefteq s_i$ and $\delta_j \trianglelefteq s_j$. It follows that $\delta_i \trianglelefteq \beta_i \triangleleft \alpha_i \trianglelefteq \gamma_i$ and $\delta_j \trianglelefteq \alpha_j \triangleleft \beta_j \trianglelefteq \gamma_j$. Thus, δ is dominated by γ on components i and j, contradicting to Lemma 7.2 part ii.

Case 2: Suppose that $l \in \{i, j\}$. Suppose w.l.o.g. that $l = i$. It follows that there exists $\delta \in \pi^{\rho}_{W+}$ such that $\delta_i \rhd \alpha_i$. Since $\alpha_i = \gamma_i$, we already have $\delta_i \rhd \gamma_i$. Next, we show that $\delta_{k^*} \rhd \gamma_{k^*}$, and thus obtain a contradiction to Lemma 7.2 part ii. Since $\delta_i \rhd \alpha_i \rhd \beta_i$, by Lemma 7.2 part ii, for each $s \in \{\alpha, \beta\}$, we have $\delta_j \unlhd s_j$ and $\delta_k \unlhd s_k$ for each $k \in N \backslash \{i, j\}$. It follows that $\delta_j \unlhd \alpha_j \lhd \beta_j$. Since $\delta_i \rhd \beta_i$ and $\delta_j \lhd \beta_j$, by Lemma 7.2 part i, $\delta_k = \beta_k$ for each $k \in N \setminus \{i, j\}$. Since $\beta_{k^*} \rhd \gamma_{k^*}$, it follows that $\delta_{k^*} \rhd \gamma_{k^*}$.

Proof of Proposition 3 Since the if part is clear, we proceed with the only if part. Suppose that II fails to hold. Then, we show that I holds. If each distinct $\alpha, \beta \in \pi^{\rho}_{W+}$ differ on a single component, then II holds. Since we suppose that this is not the case, let $\alpha, \beta \in \pi^{\rho}_{W+}$ such that $\alpha_i \rhd \beta_i$ and $\beta_j \rhd \alpha_j$. Then, let γ and k^* be as described in Lemma 7.3. Suppose w.l.o.g that $\gamma_{k^*} \rhd \alpha_{k^*}$, and thus α and β are both 1-mistake away from $\bar{s}^p$. It follows that $\alpha_i = \bar{s}^{\rho}_i$ and $\beta_j = \bar{s}^{\rho}_j$.

Now, let $\theta \in \pi^{\rho}_{W+}$. If θ differs from α on two components, then since α is 1-mistake away from $\bar{s}^p$, it follows from Lemma 7.3 that θ is also 1-mistake away from $\bar{s}^p$. If θ differs from α on a single component $l \in N$, then there are three cases.

Case 1: Suppose that $l \in N \setminus \{i, j\}$. Then, we have $\alpha_l = \beta_l = \bar{s}^{\rho}_l \rhd \theta_l$, and by Lemma 7.2 part ii, $\theta_j \unrhd \beta_j = \bar{s}^{\rho}_j \rhd \alpha_j$. It follows that θ differs from α on components j and l, contradicting that θ differs from α on a single component.

Case 2: Suppose that $l = j$. Then, since $\theta_k = \alpha_k = \bar{s}^{\rho}_k$ for each $k \in N \setminus \{j\}$, it directly follows that θ is 1-mistake away from $\bar{s}^p$.

Case 3: Suppose that $l = i$. Then, consider γ. Since $\gamma_{k^*} \lhd \beta_{k^*} = \alpha_{k^*}$, by Lemma 7.2 part ii, for each $s \in \{\beta, \alpha\}$, it must be that $\gamma_i \unrhd s_i$ and $\gamma_j \unrhd s_j$. Therefore, $\gamma_i \unrhd \alpha_i \rhd \beta_i$ and $\gamma_j \unrhd \beta_j \rhd \alpha_j$. Then, since $\alpha_i = \bar{s}^{\rho}_i$ and $\theta_i \neq \alpha_i$, we have $\gamma_i \rhd \theta_i$. Moreover, since $\theta_j = \alpha_j$, we have $\gamma_j \rhd \theta_j$. Therefore, θ is dominated by γ on components i and j, contradicting to Lemma 7.2 part ii. Thus, we conclude that $l \neq i$.

References

Agresti, A. (1984), *Analysis of ordinal categorical data*, Wiley, New York. 6

Alsina, C., Nelsen, R. B. & Schweizer, B. (1993), 'On the characterization of a class of binary operations on distribution functions', *Statistics & Probability Letters* **17**(2), 85–89. 17

Amemiya, T. (1981), 'Qualitative response models: A survey', *Journal of Economic Literature* **19**(4), 1483–1536. 6

Apesteguia, J. & Ballester, M. A. (2023*a*), 'Random utility models with ordered types and domains', *Journal of Economic Theory* **211**, 105674. 6

Apesteguia, J. & Ballester, M. A. (2023*b*), Testable implications of parametric assumptions in ordered random utility models. Working paper. 6

Apesteguia, J., Ballester, M. A. & Lu, J. (2017), 'Single-crossing random utility models', *Econometrica* **85**(2), 661–674. 3, 6, 16

Barseghyan, L., Molinari, F. & Thirkettle, M. (2021), 'Discrete choice under risk with limited consideration', *American Economic Review* **111**(6), 1972–2006. 6

Bastianello, L. & LiCalzi, M. (2019), 'The probability to reach an agreement as a foundation for axiomatic bargaining', *Econometrica* **87**(3), 837–865. 6

Caliari, D. & Petri, H. (2024), 'Irrational random utility models', *arXiv preprint arXiv:2403.10208* . 9

Dall'Aglio, G. (1972), Fréchet classes and compatibility of distribution functions, *in* 'Symposia mathematica', Vol. 9, pp. 131–150. 6, 21

Falmagne, J.-C. (1978), 'A representation theorem for finite random scale systems', *Journal of Mathematical Psychology* **18**(1), 52–72. 9

Fan, Y. & Patton, A. J. (2014), 'Copulas in econometrics', *Annual Review of Economics* **6**(1), 179–200. 6

Filiz-Ozbay, E. & Masatlioglu, Y. (2023), 'Progressive random choice', *Journal of Political Economy* **131**(3), 716–750. 3, 6, 11, 13

Fishburn, P. C. (1998), Stochastic utility, *in* S. Barbera, P. Hammond & C. Seidl, eds, 'Handbook of Utility Theory', Kluwer, Dordrecht, pp. 273–318. 9

Fréchet, M. (1935), 'Généralisation du théoreme des probabilités totales', *Fundamenta mathematicae* **25**(1), 379–387. 6, 11

Fréchet, M. (1951), 'Sur les tableaux de corrélation dont les marges sont données', *Ann. Univ. Lyon, 3ˆ e serie, Sciences, Sect. A* **14**, 53–77. 6, 11

Fréchet, M. (1958), 'Notes about the previous note', *CR Acad. Sci. Paris* **246**, 2719–2720. 21

Frick, M., Iijima, R. & Ishii, Y. (2022), 'Dispersed behavior and perceptions in assortative societies', *American Economic Review* **112**(9), 3063–3105. 6

Gresik, T. A. (2011), 'The effects of statistically dependent values on equilibrium strategies of bilateral k-double auctions', *Games and Economic Behavior* **72**(1), 139–148. 6

Hoeffding, W. (1940), 'Maßstabinvariante korrelationstheorie', *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin* **5**(3), 181–233. 6, 11

Lauzier, J.-G., Lin, L. & Wang, R. (2023), 'Pairwise counter-monotonicity', *Insurance: Mathematics and Economics* **111**, 279–287. 6

Masatlioglu, Y. & Vu, T. P. (2024), 'Growing attention', *Journal of Economic Theory* **222**, 105929. 6

Masatlioglu, Y. & Yildiz, K. (2025), 'Bounds for boundedly rational choice', *mimeo* . 6

Nelsen, R. B. (2006), *An introduction to copulas*, Springer. 6, 8, 21, 22

Nelsen, R. B., Quesada-Molina, J. J., Schweizer, B. & Sempi, C. (1996), 'Derivability of some operations on distribution functions', *Lecture Notes-Monograph Series* pp. 233–243. 17

Petri, H. (2023), 'Binary single-crossing random utility models', *Games and Economic Behavior* **138**, 311–320. 6

Schweizer, B. & Sklar, A. (2005), *Probabilistic metric spaces*, Dover. 6

Sklar, M. (1959), Fonctions de répartition à n dimensions et leurs marges, *in* 'Annales de l'ISUP', Vol. 8, pp. 229–231. 4, 6, 8

Small, K. A. (1987), 'A discrete choice model for ordered alternatives', *Econometrica: Journal of the Econometric Society* pp. 409–424. 6

Suleymanov, E. (2024), 'Branching-independent random utility model', *Journal of Economic Theory* **220**, 105880. 9

Tserenjigmid, G. (2021), 'The order-dependent Luce model', *Management Science* **67**(11), 6915–6933. 6

Turansick, C. (2022), 'Identification in the random utility model', *Journal of Economic Theory* **203**, 105489. 6

Yildiz, K. (2022), 'Self-progressive choice models', *arXiv preprint arXiv:2212.13449* . 6, 17