Ordered Probabilistic Choice*

Christopher P. Chambers† Yusufcan Masatlioglu‡ Kemal Yildiz§

January 23, 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. The connection highlights *extremal representations*–associated with the Fréchet-Hoeffding bounds–and their distinctive structural properties. Thus, we derive identification methods to uniquely determine the specific heterogeneous choice types and their corresponding weights. The unified framework elucidates the uniqueness of known representations while showcasing the potential of copula-based methods to uncover new choice models and results. Our analysis provides tools for inferring 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.

†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.

1 Introduction

Discrete choice modeling plays a crucial role in analyzing aggregate behaviors such as grocery purchasing, financial investment, job search, labor force participation, environmental conservation, and energy consumption. 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.

The model proposed by Apesteguia, Ballester & Lu (2017) addresses the issue of multiple RUM representations. The *single-crossing random utility model* (SCRUM) restricts preference heterogeneity while still accurately capturing individual choice behavior. They showed that uniqueness is guaranteed under three main assumptions: i) choice objects are ordered, ii) individuals are rational, and iii) individual types are ordered. Building on SCRUM, Filiz-Ozbay & Masatlioglu (2023) investigated the implications of ordered types for boundedly rational agents to capture factors like limited attention, willpower, shortlisting constraints, loss aversion, and pro-social behavior. The resulting model, called *progressive random choice* (PRC), accommodates any probabilistic choice behavior while retaining a unique representation.

These findings suggest that it is the ordering of types, rather than rationality, that forms the foundation of the uniqueness result. This naturally raises the question: could other plausible heterogeneity restrictions also lead to unique representations? Furthermore, since SCRUM identifies a specific RUM representation as the unique characterization of the data, what sets SCRUM's representation apart from other representations? What distinct properties does it have compared to other RUM representations? Our aim is to provide a novel framework to address these fundamental questions.

Our contribution is twofold. First, it turns out 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. To explore this, we first connect 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*.

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.

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

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 equivalence 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 plausible and intriguing probabilistic choice models, we introduce the *1-mistake model* identified by the Fréchet-Hoeffding lower bound. 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 limitations or the use of various decision-making heuristics. In an 1-mistake model, agents 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 possesses empirical content characterized by a single axiom. Unlike RUM, the 1-mistake model

[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.

provides an interpretation that avoids 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). 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, which enables researchers to investigate these models more rigorously.

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 our 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. Of course, copulas find heavy use in econometrics and finance as well—Fan & Patton (2014) provides a systematic treatment. Copula has 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).

2 Ordered Probabilistic Choice and Copulas

2.1 Preliminaries

Let X be a finite set of **alternatives**. We consider scenarios where the alternatives have 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 $\{S_i\}_{i \in N}$ be a family of **choice sets**, where the cardinality of N is denoted by n. 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 choices types: $s \unrhd s'$ if $s_i \unrhd s'_i$ for each $i \in N$.

A **probabilistic choice function** (`pcf`) ρ assigns to each choice set S_i a probability measure over S_i. We denote by $\rho(s_i, S_i)$ the probability that alternative s_i is chosen from the choice set S_i. $\mathcal{P}$ denotes 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. Let $\Omega = \{(s_i, S_i) : i \in N \text{ and } s_i \in S_i\}$. Then, the **cumulative choice function** (`ccf`) associated to ρ is $P^\rho : \Omega \to [0,1]$ such that

$$P^\rho(s_i, S_i) = \sum_{t_i \in S_i : s_i \unrhd t_i} \rho(t_i, S_i).$$

for each $(s_i, S_i) \in \Omega$. We will use P instead of P^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 $C : [0,1]^n \to \mathbb{R}$ that satisfies the following three properties. It is *grounded* if $C(u_1, u_2, \ldots, u_n) = 0$ when any $u_i = 0$, and 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.[3]

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.

Definition. Let ρ be an pcf and π be a probability distribution over choice types. Then, π **represents** ρ if for each $(x, S_i) \in \Omega$ we have

$$\rho(x, S_i) = \sum_{s : s_i = x} \pi(s).$$

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 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.

[3] This is a variant of the inclusion-exclusion principle, namely that $\sum_{T \subseteq N} C(a_C, b_{N \setminus C})| - 1|^{|C|} \geq 0$.

We next describe two new objects associated to a model $\mathcal{M}$. For each pcf ρ, let $I_{\mathcal{M}}$ 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 $\mathcal{M}$.[4]

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 $\rho \in \mathcal{M}$. This means that for any observed behavior ρ consistent with $\mathcal{M}$, there exists a single representation 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 $\rho \in$ RUM.[5]

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

[4]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 5 for an example). This distinction enables a more expressive framework for differentiation.

[5] 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 has a unique identification property: $|I_{BI-RUM}(\rho)| = 1$ for all $\rho \in \mathcal{P}_{RUM}$. Additionally, $I_{BI-RUM}(\rho)$ is always a subset of $I_{RUM}(\rho)$. Thus, while RUM and BI-RUM share the same explanatory power, they differ in their identification properties.

π is, in general, a challenging task, often requiring 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 $I : \mathcal{P}_{\mathcal{M}} \to \Delta(\mathcal{S})$ such that $I(\rho) \in I_{\mathcal{M}}(\rho)$ 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 the identification method I uniquely identifies the model $\mathcal{M}$. 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 π_C^ρ that represents ρ. For each `pcf` ρ, the associated `ccf` P^ρ specifies $P_i^\rho(s_i) = \rho(\{t_i \in S_i : s_i \trianglerighteq t_i\}, S_i)$ for each $(s_i, S_i) \in \Omega$. Then, for a given copula C and `pcf` ρ, we define the probability measure $\pi_C^\rho \in \Delta(\mathcal{S})$ where $\pi_C^\rho(s)$ is the weight assigned to the choice type $s \in \mathcal{S}$ such that the following identity holds.

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

Equation (1) defines the distribution π_C^ρ 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. Consequently, each copula C defines an identification method for $\mathcal{P}$ such that $I_C(\rho) = \pi_C^\rho$ for every `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.

2.3 Fréchet-Hoeffding Bounds

In this subsection, we provide a key discovery in copula theory, which we utilize for our purposes later. 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,

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

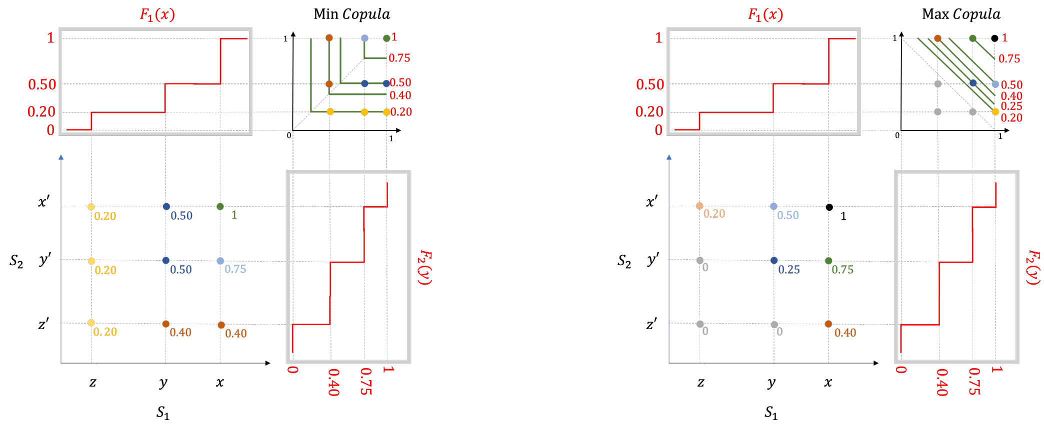
Moreover, these bounds are pointwise sharp, i.e. for each $u \in [0,1]^n$,

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

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 types of extreme dependencies. In the case of the upper bound, the two random variables are perfectly aligned, exhibiting what is 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 W and M generate two distinct representations for the same choice data. We consider two disjoint choice problems $S_1 = \{x, y, z\}$ and $S_2 = \{x', y', z'\}$ with marginal choice probabilities $\rho(z, S_1) = 0.20$, $\rho(y, S_1) = 0.30$, $\rho(z', S_2) = 0.40$, and $\rho(y', S_2) = 0.35$. We assume the reference order: $x \rhd y \rhd z$ and $x' \rhd y' \rhd z'$.

Figure 1 provides two distinct cumulative representations generated by W and M, respectively. We first calculate the cumulative marginal distributions: F_1 and F_2. We have $F_1(z) = 0.20$, $F_1(y) = 0.50$, and $F_1(x) = 1.00$. Similarly, $F_2(z') = 0.40$, $F_2(y') = 0.75$, and $F_2(x') = 1.00$. Note that F_1 and F_2 are the same cumulative marginal distributions in both panels since they are based on the choice data. In the left panel, we calculate the corresponding cumulative joint distribution using M. For example, $M(F_1(y), F_2(z')) = M(0.50, 0.40) = 0.40$ and $M(F_1(z), F_2(x')) = M(0.20, 1.00) = 0.20$. Hence, the cumulative probabilities of the types below $[y, z']$ and $[z, x']$ are 0.40 and 0.20, respectively. On the other hand, $W(F_1(y), F_2(z')) = W(0.50, 0.40) = 0$ and $W(F_1(z), F_2(x')) = W(0.20, 1.00) = 0.20$, displayed on the right panel. Then the cumulative probabilities of the types below $[y, z']$ is zero. This implies that the probability of the type $[z, z']$ is also zero.



(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 $\rho(z, S_1) = 0.20$, $\rho(y, S_1) = 0.30$, $\rho(z', S_2) = 0.40$, and $\rho(y', S_2) = 0.35$. 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.

Figure 2 provides the unique weights associated with each type according to W and M. This figure is based on the cumulative distributions provided in Figure 1. Since $M(F_1(z), F_2(z')) = M(F_1(z), F_2(x')) = 0.20$, 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, the types in the support of W are ordered in a decreasing sequence. Additionally, while M assigns a weight of 0.25 to the highest type $[x, x']$, W assigns it a weight of zero.

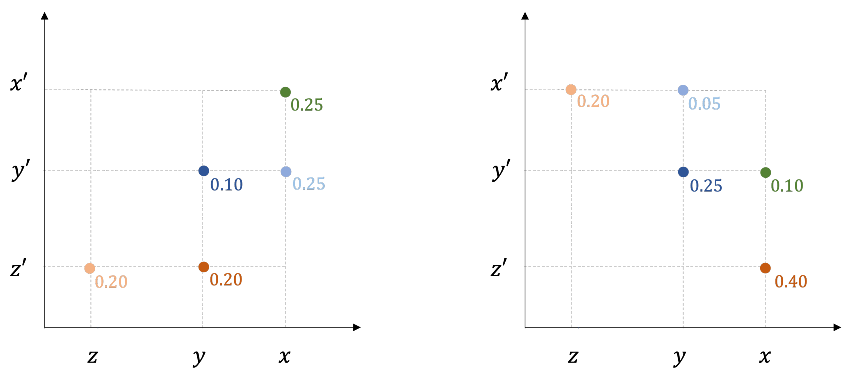


Figure 2: The representation of the distribution over deterministic choice functions for Fréchet-Hoeffding bounds shows two panels. The left panel is based on the upper bound, illustrating the distribution using the min-copula. The right panel depicts the weights induced by the lower bound.

3 Progressive random choice and the FH-upper bound

In this section, we show that the min-copula M is the identification method for the model of Filiz-Ozbay & Masatlioglu (2023) (FM). Then, 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 environmental 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 $\{s^1, \ldots, s^T\}$ is **progressive** with respect to $\rhd$ if $s_i^t \unrhd s_i^{t+1}$ for each $i \in N$ and $t \in \{1, \ldots, T-1\}$. 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,

$$PRC := \{\langle \rho, \pi \rangle \, | \pi \text{ represents } \rho \text{ and the support of } \pi \text{ is progressive}\}.$$

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 one can establish their results by utilizing 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, $I_M(\rho)$ 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 PRC: For a given choice type s, the PRC representation assigns a higher probability to choice types dominating s compared to any other probability distribution π over choice types that generates 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 $s \in S$, the PRC representation of ρ, denoted by π^ρ_{PRC}, satisfies the inequalities*

$$\sum_{t \in \mathcal{S}: t \unrhd s} \pi^\rho_{PRC}(t) \geq \sum_{t \in \mathcal{S}: t \unrhd s} \pi(t) \;\; \textit{and} \;\; \sum_{t \in \mathcal{S}: s \unrhd t} \pi^\rho_{PRC}(t) \geq \sum_{t \in \mathcal{S}: s \unrhd t} \pi(t).$$

Proof. For the second inequality, since the PRC representation corresponds to the min-copula, it is sufficient to show that $\pi(\{s' : s \unrhd s'\}) \leq \min\{P^\rho_1(s_1), \dots, P^\rho_n(s_n)\}$. But for any i, $\pi(\{s' : s \unrhd s'\}) \leq \pi(\{s' : s_i \unrhd s'_i\}) = P^{\rho_i}(s_i)$. So this establishes the result.

To show the statement about $\pi(\{s' : s' \unrhd s\})$, (the first inequality), we first introduce a piece of notation. For $i \in N$ and s_i, we let $P^\rho_i(s_i - 1) = \sum_{t_i \in S_i : s_i \rhd t_i} \rho(t_i, S_i)$. So, if s_i is the $\rhd$ initial element $P^\rho_i(s_i - 1) = 0$, for example.

Then according to the min-copula, we see that for PRC, the probability of $\{s' : s' \trianglerighteq s\}$ is given by $1 - \sum_{\varnothing \neq C \subseteq N} \min_{i \in C}\{P_i^\rho(s_i - 1)\}|-1|^{|C|}$, by inclusion-exclusion. Without loss, let us assume temporarily that $i \leq j$ implies $P_i^\rho(s_i - 1) \leq P_j^\rho(s_j - 1)$. We rewrite the sum as: $1 - \sum_{i=1}^n \sum_{C: i \in C \subseteq N \setminus \{1,\ldots,i-1\}} 1 - \{P_i^\rho(s_i - 1)\}|-1|^{|C|}$, where $\{1, \ldots, i-1\} = \varnothing$ when $i = 1$. Observe then that for each $i < n$, $\sum_{C: i \in C \subseteq N \setminus \{1,\ldots,i-1\}} |-1|^{|C|} = \sum_{k=0}^{n-i} \binom{n-i}{k} |-1|^{1+k} = (-1) \sum_{k=0}^{n-i} \binom{n-i}{k} |-1|^k = (-1)(1-1)^{n-i} = 0$ by the binomial formula. On the other hand for $i = n$, the only C in the sum is $C = \{n\}$ and so in this case $\sum_{C: n \in C \subseteq N \setminus \{1,\ldots,n-1\}} |-1|^{|C|} = -1$ (this also follows from the binomial formula). So overall the expression evaluates as $1 - P_n^\rho(s_n - 1)$. The same argument establishes that the probability given to $\{s' : s' \trianglerighteq s\}$ is always $1 - \max_{i \in N} P_i^\rho(s_i - 1) = \min_{i \in N}(1 - P_i^\rho(s_i - 1))$ even absent the assumption that $i \leq j$ implies $P_i^\rho(s_i - 1) \leq P_j^\rho(s_j - 1)$. We now observe that for any π, $\pi(\{s' : s' \trianglerighteq s\}) \leq \pi(\{s' : s'_i \trianglerighteq s_i\}) = 1 - P_i^\rho(s_i - 1)$, again establishing the result. ■

An immediate implication is that the probability assigned to the choice type maximizing the underlying reference order is maximized by PRC. Similarly, PRC also maximizes the probability assigned to the choice type minimizing the underlying reference order. To get an intuition, recall that the PRC implies has the property that for all s, s' in the support of π, we have either $s \trianglerighteq s'$ or $s' \trianglerighteq s$. Proposition 1 strengthens this intuition to require that for *any* s, the probability of obtaining an s' comparable to it (in either direction) according to $\trianglerighteq$ is maximized.

As noted in the introduction, SCRUM identifies a specific RUM representation as its unique form. Let us formally define SCRUM as follows:

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

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) shows that data satisfies both *centrality* and *regularity* if and only if it has a SCRUM representation.[6] This implies that the min-copula assigns positive

[6] *Regularity*: If $B \subset A$, 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 $\rho(z, \{z, y\}) = \rho(z, \{x, y, z\})$.

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.

Finally, the min-copula acts as an identification method for any model that incorporates a progressive structure along with certain additional support conditions. 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 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. Overall, copula theory helps us link plausible restrictions on underlying heterogeneity to a functional form that identifies unique weights.

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 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.[7] While every copula is a quasi-copula, there exist quasi-copulas that are not copulas. Now, we define when an `pcf` is identified by a quasi-copula.

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

$$\pi_Q^\rho(\{s' : s_i \trianglerighteq s'_i \text{ for each } i \in N\}) = Q(P_1^\rho(s_1), \ldots, P_n^\rho(s_n)). \tag{2}$$

As with copulas, the corresponding representation is unique when it exists. Let $\mathcal{P}_Q$ denote the set of all `pcf`s identified by Q. Note that if Q is a copula, then $\mathcal{P}_Q = \mathcal{P}$. That is why we have $\mathcal{P} = \mathcal{P}_{PRC} = \mathcal{P}_M$. Given a quasi-copula Q, each pair $\langle \rho, \pi_Q^\rho \rangle$ where Q identifies ρ constitutes a model. Formally, a quasi-copula Q induces a model

$$\mathcal{M}_Q := \{\langle \rho, \pi_Q^\rho \rangle \mid \pi_Q^\rho \text{ is a probability measure and represents } \rho\}.$$

We interpret an `pcf` for which the application of Q does not result in a probability distribution as being *ruled out* by $\mathcal{M}_Q$ (or simply by Q). Note that $\mathcal{M}_Q$ is uniquely identified and Q is always the identification method for $\mathcal{M}_Q$.

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

[7] C satisfies the Lipschitz condition if $|C(\mathbf{u}) - C(\mathbf{u}')| \leq \sum_i |u_i - u'_i|$ for every $\mathbf{u}$ and $\mathbf{u}'$ in $[0,1]^n$.

that each type is either entirely rational (free of mistakes) or makes a mistake in a single-choice set. Formally, a choice type c is **near $\rhd$-optimal** if there exists at most one choice set S_i such that $c(S_i)$ differs from the $\rhd$-best element in S_i. Let $\mathcal{N}_{\rhd}$ denote all **near $\rhd$-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 $x \rhd y \rhd z$. 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 $\rhd$-optimal choice types. Now we can define the 1-mistake model. An `pcf` ρ has a 1**-mistake** representation with respect to $\rhd$ if there exits a probability distribution π over near $\rhd$-optimal choice functions that generates ρ, i.e., 1**-mistake** $:= \{\langle \rho, \pi \rangle \mid \pi \in \Delta(\mathcal{N}_{\rhd})$ 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 (describes $\mathcal{P}_{\textbf{1-mistake}}$). 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 $\rhd$-best element in S_i. Then,*

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

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 $\sum_{i \in N}(1 - \rho(\bar{s}_i, S_i)) \leq 1$. Conversely, let ρ be an `RCF` such that $\sum_{i \in N}(1 - \rho(\bar{s}_i, S_i)) \leq 1$. 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 $\pi(s^i) = \rho(s_i, S_i)$ and $\pi(\bar{s}) = 1 - \sum_{i \in N}(1 - \rho(\bar{s}_i, S_i))$, which is nonnegative by our assumption. Thus, π generates ρ, and has a support consisting of near $\trianglerighteq$-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, $\max\{\sum_{i=1}^{n} P^{\rho}_i(\bar{s}_i) + 1 -$

$n, 0\} = \max\{1, 0\} = 1$. 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 $>$-worse than s_i. Recall that by Part i, $\sum_{i=1}^n (1-\rho(\bar{s}_i, S_i)) \leq 1$. 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 $S_i^+ = \{s_i \in S_i : p(s_i, S_i) > 0\}$ and $\bar{s}_i^\rho$ ($\underline{s}_i^\rho$) be the $\triangleright$-best(worst) element in S_i^+. Let $\bar{s}^p = [\bar{s}_1^\rho, \ldots, \bar{s}_n^\rho]$ and $\underline{s}^p = [\underline{s}_1^\rho, \ldots, \underline{s}_n^\rho]$.

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 $\underline{s}^p$ that generates ρ, or*

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

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 based on the underlying ordering.

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 here. Instead, we provide several interesting copulas, which could be useful for generating and identifying new models of ordered probabilistic choice.

Example 1 (Independent Copula)**.** The Independent copula is the copula that results from a dependency structure in which each individual variable is independent of each other. Probably it is the simplest and most straightforward copula, where:

$$\Pi(u_1, u_2, \cdots, u_n) := \Pi_i \, u_i$$

The independent copula is independent of the reference order. Independent copulas are used in various fields, including statistical modeling, finance, and machine learning.

Example 2 (Fréchet Copula Family)**.** Suppose that $\{C_\alpha\}$ is a one-parameter family of quasi-copulas, which is a version of Fréchet copula family (Fréchet (1958)). C_α is a linear combination of the FH-lower and the FH-upper bounds such that $C_0 = W$ and $C_1 = M$. C_α is not a copula in general. For $\alpha \in (0, 1)$, the support of C_α consists of the union of progressive and near-optimal choice types.

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

As W, one can find conditions on the marginal distribution such that C_α delivers a joint distribution. As we know from the last section, the condition in Axiom 1 makes W delivering non-negative weights. Since C_α is a combination of W and M, the condition for C_α

should be weaker than the condition for W. Indeed, the condition must be monotonic with respect to α.

Example 3 (Threshold Copula Family)**.** This family is based on an example in Nelsen (2006). C_t is a mixture of W and M given a threshold level t. While C_t behaves according to M for lower probabilities, it acts as W for higher probabilities. Formally,

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

Similar to Example 2, C_t is a combination of W and M such that $C_0 = W$ and $C_1 = M$. Again, C_t is not a copula in general. The support of C_t could be outside of the union of progressive and near-optimal choice types.

One can also construct new copulas by using existing copulas. A straight approach to constructing new multidimensional copulas would be to use 2-copulas to join or couple other 2-copulas, as the following examples illustrate:

Example 4. New quasi-copulas can also be constructed by grouping choice problems into clusters and applying distinct aggregation functions for within-group and between-group interactions, such as:

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

Example 5. Suppose that there are three groups of choice problems: $G_1 = \{1, \cdots, k_1\}$, $G_2 = \{k_1 + 1, \cdots, k_2\}$, $G_3 = \{k_2 + 1, \cdots, n\}$, representing easy, medium, and high difficulty levels of choice problems, respectively. Then the function W is applied across these groups as follows. Note that this might not be a copula.

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

6 Conclusion

We demonstrated that the min-copula serves as an identification method for various models. This approach is particularly beneficial for identification, as it eliminates the need for complex constructions by providing an exact formula to calculate the unique weights assigned to choice types in the support.

Additionally, we showed that copula theory serves as a powerful tool for uncovering plausible probabilistic choice models. The 1-mistake model exemplifies such a model, combining both novelty and practical relevance. 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_W^ρ, denoted by ${\pi_W^\rho}^+$. That is, there exist no two choice types $s, s' \in {\pi_W^\rho}^+$ 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_W^ρ is an $\rhd$-antichain.*

Proof. By contradiction, suppose that there exist $s, s' \in {\pi_W^\rho}^+$ such that $s_i \rhd s'_i$ for each $i \in N$. Let F_W^ρ be the `CDF` associated with W and ρ. That is, $F_W^\rho(s) = W(P_1^\rho(s_1), \ldots, P_n^\rho(s_n))$ for each $s \in S$ (equivalently $F_W^\rho(s) = \sum_{t: s \trianglerighteq t} \pi_W^\rho(t)$). Now, let $[s-1]$ be the element of S such that $[s-1]_i = s_i - 1$ for each $i \in N$.[8] Then,

$$\pi_W^\rho(s) = \sum_{M \subseteq N} (-1)^{|M|} \, F_W^\rho([s-1]Ms). \tag{3}$$

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

$$\pi_W^\rho(s) = \sum_{M \subseteq N} (-1)^{|M|} \left(\sum_{i \in M} P_i^\rho(s_i) + \sum_{i \in N \setminus M} P_i^\rho(s_i - 1) - n + 1 \right) \tag{4}$$

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 ${\pi_W^\rho}^+$ 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_W^ρ, 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_W^ρ with π and F_W^ρ with F. Then, since π generates ρ, we can rewrite $F(s) = 1 - n + \sum_{i \in N} P_i^\rho(s_i)$ 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 $t^* \in {\pi_W^\rho}^+$ with $s \triangleright t^*$, we get

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

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 $\alpha, \beta \in \pi_W^{\rho+}$ such that $\alpha_i > \beta_i$ and $\beta_j > \alpha_j$. Then, there exists $\gamma \in \pi_W^{\rho+}$ and $k^* \in N$ such that*

1. *$\gamma_{k^*} \neq \alpha_{k^*} = \beta_{k^*}$, and*
2. *If $\alpha_{k^*} > \gamma_{k^*}$ then α and β are 1-mistake away from $\bar{s}^p$; if $\alpha_{k^*} < \gamma_{k^*}$ 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 $p(\alpha_{k^*}, S_{k^*}) < 1$. It follows that there exists $\gamma \in \pi_W^{\rho+}$ such that $\gamma_{k^*} \neq \alpha_{k^*} = \beta_{k^*}$. Thus, 1 holds. Next, suppose w.l.o.g. that $\alpha_{k^*} > \gamma_{k^*}$. Then, by Lemma 7.2 part ii, for each $s \in \{\beta, \alpha\}$, it must be that $\gamma_i \geq s_i$ and $\gamma_j \geq s_j$. Therefore, $\gamma_i \geq \alpha_i > \beta_i$ and $\gamma_j \geq \beta_j > \alpha_j$. 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 $\gamma_{k^*} < \beta_{k^*} = \alpha_{k^*}$, we would be showing that α and β are 1-mistake away from $\underline{s}^p$). That is, we claim that $\alpha_i = \bar{s}_i^\rho$, $\beta_j = \bar{s}_j^\rho$, and $\bar{s}_k^\rho = \alpha_k = \beta_k$ 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 $\delta \in \pi_W^{\rho+}$ such that $\delta_l > \beta_l = \alpha_l$. By Lemma 7.2 part ii, for each $s \in \{\alpha, \beta\}$, it must be that $\delta_i \leq s_i$ and $\delta_j \leq s_j$. It follows that $\delta_i \leq \beta_i < \alpha_i \leq \gamma_i$ and $\delta_j \leq \alpha_j < \beta_j \leq \gamma_j$. 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 $\delta \in \pi_W^{\rho+}$ such that $\delta_i > \alpha_i$. Since $\alpha_i = \gamma_i$, we already have $\delta_i > \gamma_i$. Next, we show that $\delta_{k^*} > \gamma_{k^*}$, and thus obtain a contradiction to Lemma 7.2 part ii. Since $\delta_i > \alpha_i > \beta_i$, by Lemma 7.2 part ii, for each $s \in \{\alpha, \beta\}$, we have $\delta_j \leq s_j$ and $\delta_k \leq s_k$ for each $k \in N \setminus \{i, j\}$. It follows that $\delta_j \leq \alpha_j < \beta_j$. Since $\delta_i > \beta_i$ and $\delta_j < \beta_j$, by Lemma 7.2 part i, $\delta_k = \beta_k$ for each $k \in N \setminus \{i, j\}$. Since $\beta_{k^*} > \gamma_{k^*}$, it follows that $\delta_{k^*} > \gamma_{k^*}$. ■

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

Now, let $\theta \in {\pi_W^\rho}^+$. 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 $\alpha_l = \beta_l = \bar{s}_l^\rho > \theta_l$, and by Lemma 7.2 part ii, $\theta_j \geq \beta_j = \bar{s}_j^\rho > \alpha_j$. 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 $\theta_k = \alpha_k = \bar{s}_k^\rho$ 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 $\gamma_{k^*} < \beta_{k^*} = \alpha_{k^*}$, by Lemma 7.2 part ii, for each $s \in \{\beta, \alpha\}$, it must be that $\gamma_i \geq s_i$ and $\gamma_j \geq s_j$. Therefore, $\gamma_i \geq \alpha_i > \beta_i$ and $\gamma_j \geq \beta_j > \alpha_j$. Then, since $\alpha_i = \bar{s}_i^\rho$ and $\theta_i \neq \alpha_i$, we have $\gamma_i > \theta_i$. Moreover, since $\theta_j = \alpha_j$, we have $\gamma_j > \theta_j$. 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. 5

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. 15

Amemiya, T. (1981), 'Qualitative response models: A survey', *Journal of economic literature* **19**(4), 1483–1536. 5

Apesteguia, J. & Ballester, M. A. (2023*a*), 'Random utility models with ordered types and domains', *Journal of Economic Theory* **211**, 105674. 5

Apesteguia, J. & Ballester, M. A. (2023*b*), Testable implications of parametric assumptions in ordered random utility models. Working paper. 5

Apesteguia, J., Ballester, M. A. & Lu, J. (2017), 'Single-crossing random utility models', *Econometrica* **85**(2), 661–674. 2, 5, 14

Barseghyan, L., Molinari, F. & Thirkettle, M. (2021), 'Discrete choice under risk with limited consideration', *American Economic Review* **111**(6), 1972–2006. 5

Bastianello, L. & LiCalzi, M. (2019), 'The probability to reach an agreement as a foundation for axiomatic bargaining', *Econometrica* **87**(3), 837–865. 5

Dall'Aglio, G. (1972), Fréchet classes and compatibility of distribution functions, *in* 'Symposia mathematica', Vol. 9, pp. 131–150. 5, 19

Falmagne, J.-C. (1978), 'A representation theorem for finite random scale systems', *Journal of Mathematical Psychology* **18**(1), 52–72. 8

Fan, Y. & Patton, A. J. (2014), 'Copulas in econometrics', *Annual Review of Economics* **6**(1), 179–200. 5

Filiz-Ozbay, E. & Masatlioglu, Y. (2023), 'Progressive random choice', *Journal of Political Economy* **131**(3), 716–750. 2, 5, 12

Fishburn, P. C. (1998), Stochastic utility, *in* S. Barbera, P. Hammond & C. Seidl, eds, 'Handbook of Utility Theory', Kluwer, Dordrecht, pp. 273–318. 8

Fréchet, M. (1935), 'Généralisation du théoreme des probabilités totales', *Fundamenta mathematicae* **25**(1), 379–387. 5, 9

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. 5, 9

Fréchet, M. (1958), 'Notes about the previous note', *CR Acad. Sci. Paris* **246**, 2719–2720. 19

Frick, M., Iijima, R. & Ishii, Y. (2022), 'Dispersed behavior and perceptions in assortative societies', *American Economic Review* **112**(9), 3063–3105. 5

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. 5

Hoeffding, W. (1940), 'Maßstabinvariante korrelationstheorie', *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin* **5**(3), 181–233. 5, 9

Lauzier, J.-G., Lin, L. & Wang, R. (2023), 'Pairwise counter-monotonicity', *Insurance: Mathematics and Economics* **111**, 279–287. 5

Masatlioglu, Y. & Vu, T. P. (2024), 'Growing attention', *Journal of Economic Theory* **222**, 105929. 5

Nelsen, R. B. (2006), *An introduction to copulas*, Springer. 5, 7, 19, 20

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. 15

Petri, H. (2023), 'Binary single-crossing random utility models', *Games and Economic Behavior* **138**, 311–320. 5

Schweizer, B. & Sklar, A. (2005), *Probabilistic metric spaces*, Dover. 5

Sklar, M. (1959), Fonctions de répartition à n dimensions et leurs marges, *in* 'Annales de l'ISUP', Vol. 8, pp. 229–231. 3, 5, 7

Small, K. A. (1987), 'A discrete choice model for ordered alternatives', *Econometrica: Journal of the Econometric Society* pp. 409–424. 5

Suleymanov, E. (2024), 'Branching-independent random utility model', *Journal of Economic Theory* **220**, 105880. 8

Tserenjigmid, G. (2021), 'The order-dependent luce model', *Management Science* **67**(11), 6915–6933. 5

Turansick, C. (2022), 'Identification in the random utility model', *Journal of Economic Theory* **203**, 105489. 5

Yildiz, K. (2022), 'Self-progressive choice models', *arXiv preprint arXiv:2212.13449* . 5