THEMA Working Paper n°2023-06
CY Cergy Paris Université, France

Best, worst, and Best&worst choice probabilities for logit and reverse logit models

André de Palma, Karim Kilani

April 2023

Best, worst, and Best&worst choice probabilities for logit and reverse logit models

André de Palma* and Karim Kilani†

April 2023

Abstract

This paper builds upon the work of Professor Marley, who, since the beginning of his long research career, has proposed rigorous axiomatics in the area of probabilistic choice models. Our study concentrates on models that can be applied to best and worst choice scaling experiments. We focus on those among these models that are based on strong assumptions about the underlying ranking of the alternatives with which the individual is assumed to be endowed when making the choice. Taking advantage of an inclusion-exclusion identity that we showed a few years ago, we propose a variety of best-worst choice probability models that could be implemented in software packages that are flourishing in this field.

Keywords. Best-worst scaling experiments; Logit model; Random utility models; Reverse logit model **JEL classification.** C25; C35;

*CY Cergy Paris Université, THEMA, Cergy-Pontoise, France, andre.de-palma@cyu.fr

†**Corresponding author.** LIRSA, Conseratoire National des Arts et Métiers, 2 rue Conté, 75003, Paris, France, karkil2205@gmail.com

1 Introduction

Professor Tony Marley, to whom we pay a well-deserved tribute in this special issue of this journal, is undoubtedly the pioneer of probabilistic choice models in which we are interested in the best choice, the worst choice, and/or both the best and the worst choice. His theoretical contributions on the subject can be traced back to Marley [1968] which can be considered as the seminal paper on this topic.

This first theoretical work has gained in maturity due to Marley's collaboration with Jordan Louviere who had seen very early and in an independent way, the interest of best-worst scaling experiments. Their joint work Marley and Louviere [2005] published in the Journal of Mathematical Psychology set new milestones in this area.

Since then, the number of works on Best-worst scaling experiments has not stopped growing. Goggle Scholar currently lists more than six hundred published papers on the subject in areas as diverse as Health, Marketing, Transportation, etc. In 2015, their joint book Louviere et al. [2015] published with Terry Flynn as a third author, makes a major synthesis of theoretical and empirical contributions of the moment.

Meanwhile, a number of packages of *R* or *Python* have emerged that allow users to implement and estimate such models. The most remarkable and comprehensive is probably the Apollo package [see Hess and Palma, 2019]. The majority of the packages implement the MaxDiff model that constructs the probability of the simultaneous best and worst choices as the product of a logit best choice probability in the whole choice set, multiplied by a probability, of the same logit type but with negative sign scales, which represents the worst choice probability in the set to the exclusion of the best choice. We propose here other models almost as simple, but with different theoretical foundations.

In this article, we rely in particular on the research of Marley and Louviere [2005]. Our objective is to revisit the foundations of these models, with reference to the theory of choice developed by Luce [1959]. Our objective is to better position these models, by contributing to the enrichment of the existing literature on the subject, in particular with respect to the classification and rationalization of these models.

In this paper, we focus only on models where the individual is supposed to have a strict underlying ranking of the alternatives, but represented in a probabilistic way by the modeler. This is the approach we took in de Palma et al. [2017], even if we had a slightly stronger assumption than here of a Random Utility Model [RUM, see Anderson et al., 1992, McFadden, 1981, 2001], which is not really necessary.

We provide here a new general and rather simple proof of an inclusion-exclusion type identity which seems to us fundamental and which would benefit from being better known. This would open up many new perspectives for models designed to analyze best and worst choice experiments. Indeed, we have shown that the probabilities of the best and worst choices are totally linked by this form of inclusion-exclusion identity. Any worst choice probability can be obtained from the best choice probabilities by an alternating identity, and vice versa. The content of the rest of the paper, which will highlight this modeling, is described in the paragraphs that immediately follow.

In Section 2, we state the underlying rationality assumption required in the models we study. We then define the best and worst choice probabilities in this framework. We then formulate and prove the identity that links these two types of choice probabilities.

In Section 3, we recall the Luce's choice axiom [Luce, 1959] based on the best choice probabilities - in this case, the induced probabilities of the standard Luce model are often called the multinomial logit model (MNL). We state another axiom of the same type, this time based on worst choice probabilities. The corresponding probabilities of this alternative construct are what we will refer to as the reverse Luce model or Reverse Multinomial Logit Model (RMNL). The above identity allows us to explicitly determine the worst choice probabilities in an MNL and the best choice probabilities in an RMNL. Next, we show that a model bringing together the two axioms of the standard Luce model and the inverse Luce model produces a fully symmetric model. This is of little use from a practical point of view and represents a somewhat impossibility result.

In section 4, we address the question of the best and worst choices being considered simultaneously. We refer to their probabilities as the best&worst choice probabilities. We show that when the MNL is generated through a RUM, even if the utilities are correlated as in the model we call Gumbel-Strauss model, we obtain best&worst choice probabilities that are a product of a logit probability by an alternating sum of logit probabilities. From this, a variety of best and worst probabilities emerge, enriching the models already known for estimating experiments where the respondents simultaneously reveal their best and worst choices.

Finally, in Section 5, we summarize our results and propose theoretical and empirical research avenues to further extend the scope of best and worst scaling experiments.

2 The best and worst choice probabilities identity

The choice setting

The total master choice set, denoted by T, is assumed to be finite and containing at least three alternatives ($n \equiv |T| \geq 3$, where $|\cdot|$ denotes set cardinality). Individuals' preferences for the alternatives are assumed random and are represented in the following way: An individual selected at random from a population strictly corresponds to ranked set of alternatives. This signifies that the individual orders a number of options according to those they prefer the most, to those least preferred, and that two alternatives are never equally desired. It is also feasible that an individual who has to reveal their ranking of alternatives may modify the order in a seemingly random fashion from one experiment to another. This last point is often adopted in mathematical psychology, while the first definition is favored in econometrics.

More formally, a strict linear order, which is similar to a permutation or a ranking of the alternatives of T, is denoted using the binary relation symbol $\succ$. We must read the notation for events in the following manner:[1]

$$x \succ y \iff x \text{ is better than } y \iff y \text{ is worse than } x, \tag{1}$$

where x and y are two different alternatives belonging to T. The random nature of preferences is represented by a probability distribution $P(\cdot)$ on the universe Ω consisting of all the strict linear orders (or rankings for short) on the choice set T. Recall that the number of rankings of the alternatives of T is factorial $|T|$.

[1] A more rigorous approach for the notation of $x \succ y$ would be $\{\succ \in \Omega : x \succ y\}$. To preserve notation, we use the former expression.

The very broad hypotheses we make in this paper are specified below:

Assumption 1. *Individual preferences are represented by a probability distribution $P(\cdot)$ on the set of strict linear orders over the total set of alternatives T.*

Best and worst choice probabilities

A more general notation than that we described for two alternatives in (1), instead puts into practice the use of a subset of alternatives. Let x be any alternative of T, and Y be any nonempty subset of alternatives within T, which does not contain x, i.e. $x \notin Y$.

The following event

$$(x \succ Y) \equiv \bigcap_{y \in Y} (x \succ y), \tag{2}$$

must be understood as the subset of rankings where the alternative x is strictly preferred over all the alternatives contained within Y. The probabilities of these events are referred to as the *best choice probabilities.*

In a symmetrical manner, we can define the following event

$$(Y \succ x) \equiv \bigcap_{y \in Y} (y \succ x), \tag{3}$$

where x this time represents the worst of all the alternatives contained within Y. We refer to the probabilities of such events as the *worst choice probabilities.*

The identity relating the best and the worst choice probabilities

We show that the events in which x is considered as better or worse than other alternatives is closely related according to particular identities. First of all, let us define the best and the worst choice probabilities in the master choice set T as being respectively the following positive numbers

$$b_x \equiv P(x \succ T \setminus \{x\}); w_x \equiv P(T \setminus \{x\} \succ x), x \in T. \tag{4}$$

Note that they satisfy the following normalization identity

$$\sum_{x \in T} b_x = \sum_{x \in T} w_x = 1. \tag{5}$$

In this article we shall use the following fairly compact notations

$$b_Y \equiv \sum_{y \in Y} b_y;\ w_Y \equiv \sum_{y \in Y} w_y, \emptyset \subseteq Y \subseteq T. \tag{6}$$

By definition, we have $b_\emptyset = w_\emptyset = 0$ and $b_T = w_T = 1$.

First note that if only two alternatives x and y are involved, the event or the subset of rankings such that $(y \succ x)$ is the complement of $(x \succ y)$. By the complement rule in probability, we obtain the following identity for the binary case

$$P(y \succ x) = 1 - P(x \succ y). \tag{7}$$

Conversely, we can also interpret this simple relation as the probability that y is the best choice rewritten as a function of the probability that y is the worst choice.

Then, for three different alternatives x,y,$z \in T$, as long as the complement of the intersection is the same as the union of the complement, the subset of rankings such that $(\{y,z\} \succ x)$ is the complement of $(x \succ y) \cup (x \succ z)$. Applying both the complement and addition rules in probability, we get the following identity for the trinary case

$$P(\{y,z\} \succ x) = 1 - P(x \succ y) - P(x \succ z) + P(x \succ \{y,z\}). \tag{8}$$

The left side is the worst choice probability associated to the alternative x, while the right side contains nothing but best choice probabilities for this same alternative x. We can also rearrange and rewrite this using simple algebra and binary case identity (7). The previous identity becomes

$$P(x \succ \{y,z\}) = 1 - P(y \succ x) - P(z \succ x) + P(\{y,z\} \succ x).$$

The generalization of such binary and trinary identities is possible as demonstrated in the next theorem. We will artificially and for notational convenience extend the notations (2) and (3) to the case where Y is the empty set by setting $P(x \succ \emptyset) = P(\emptyset \succ x) \equiv 1$. We have:

Theorem 1. *Under Assumption 1, the worst choice probabilities are related to the best choice probabilities by the following identity:*

$$P(Y \succ x) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} P(x \succ Y'). \tag{9}$$

Conversely, the best choice probabilities are related to the worst choice probabilities by:

$$P(x \succ Y) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} P(Y' \succ x). \tag{10}$$

Proof. First note that $I(Y \succ x) = \prod_{y \in Y} I(y \succ x)$,where $I(\cdot)$ is the event indicator function. The above displayed expression can be rewritten as follows

$$I(Y \succ x) = \prod_{y \in Y} (1 - I(x \succ y)).$$

Expansion of the product of the above expression yields

$$I(Y \succ x) = 1 + \sum_{Y' \subseteq Y} (-1)^{|Y'|} \prod_{y \in Y'} I(x \succ y),$$

which can be rewritten as follows

$$I(Y \succ x) = 1 + \sum_{Y' \subseteq Y} (-1)^{|Y'|} I(x \succ Y').$$

Since by convention we set: $(x \succ \emptyset) \equiv \Omega$ so that $I(x \succ \emptyset) = 1$, the above displayed equation, can be rewritten in a slightly more compact form by adding the empty set to the sum of its RHS

$$I(Y \succ x) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(x \succ Y').$$

Applying the expectation operator to both sides of that equation, and by the linearity of that operator, the indicator function $I(\cdot)$ can be replaced by the probability function $P(\cdot)$, obtaining (9). The proof of (10) goes along the same lines as above by systematically switching the positions of x, Y and Y' with respect to the symbol $\succ$. $\square$

3 Models à la Luce

We introduce here some axiomatic models in the same vein as Luce [1959]. We start by the standard Luce model of which the hypotheses shed light on the best choice probabilities, and we recall the classic form of best choice probabilities induced by this axiom. Thanks to our identity outlined in Theorem 1, the form of worst choice probabilities is subsequently obtained. Next, we propose an axiomatic symmetry, that of Luce, but one which supports the worst choice probabilities called the Reverse Luce Axiom. We derive this according to the same principle as for the standard model, both for the worst choice probabilities and best choice probabilities introduced by our identity. Finally, in our last subsection, we demonstrate a model that verifies that both axioms are perfectly symmetrical and offer little interest from a practical point of view.

Luce's axiom

We start by recalling one of the fundamental results of axiomatic random preferences, that is to say, Luce's axiom.

Axiom 1 (Luce's axiom)**.** *The best choice probabilities satisfy, for any* $x,y \in T$, $x \neq y$, *and any* $Z \subset T \setminus \{x,y\}$, *the following property:*

$$\frac{P(x \succ \{y\} \cup Z)}{P(y \succ \{x\} \cup Z)} = \frac{b_x}{b_y}. \tag{11}$$

This axiom suggests that ratio between the probability that x is the best choice and probability that y is the best choice, does not depend on the presence of other alternatives. In other words, the alternatives of Z, are described as irrelevant alternatives. This ratio must, among other things, correspond with the ratio of choice probabilities when all the alternatives of the master set T are implicated in the choice. We sometimes refer to this axiom as Independence of Irrelevant Alternatives (IIA) property.

One of the fundamental results published in Luce [1959] and which is the basic of the MNL, is reiterated and demonstrated once again in the following proposition:

Theorem 2 (Luce, 1959)**.** *The best choice probabilities satisfy Luce's choice* *axiom (Axiom 1) iff the best choice probabilities have the MNL form given by:*

$$P(x \succ Y) = \frac{b_x}{b_x + b_Y}. \tag{12}$$

Proof. By the law of total probability, we have the following normalizing identity

$$P(x \succ Y) + \sum_{y \in Y} P(y \succ \{x\} \cup Y \setminus \{y\}) = 1.$$

Then, by Axiom 1, we have

$$\frac{P(y \succ \{x\} \cup Y \setminus \{y\})}{P(x \succ Y)} = \frac{P(y \succ \{x\} \cup (Y \setminus \{y\})}{P(x \succ \{y\} \cup (Y \setminus \{y\})} = \frac{b_y}{b_x},$$

implying

$$P(y \succ \{x\} \cup Y \setminus \{y\})) = \frac{b_y}{b_x} P(x \succ Y).$$

Therefore, the above normalizing identity leads to

$$P(x \succ Y) = \frac{1}{1 + b_y/b_x},$$

which can be rewritten as in (12). □

We can go on to use the previously outlined theorems so as to deduce the form of worst choice probabilities. We adopt the usual convention that an empty sum is zero, i.e. $\sum_{y \in \emptyset} b_y = 0$.

Proposition 1. *Luce's axiom (Axiom 1) holds iff the worst choice probabilities have the following form:*

$$P(Y \succ x) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} \frac{b_x}{b_x + b_{Y'}}. \tag{13}$$

Proof. Applying Identity (9) of Theorem 1 with the best , we get the worst choice probabilities formula provided in (13).

The obtained result is a necessary and sufficient condition as long as Theorem 1 gives us a relation which links the worst choice probabilities to the best choice probabilities whose form is given by Theorem 2, which itself gives necessary and sufficient conditions. $\square$

Reverse Luce's axiom

A natural extension of the previous MNL involves transferring the IIA property from the best choice probabilities to the worst choice probabilities. We call this the Reverse Luce choice axiom that we will now outline.

Axiom 2 (Reverse Luce's choice axiom)**.** *The worst choice probabilities satisfy, for any different alternatives $x,y \in T$, and any $Z \subset T \setminus \{x,y\}$, the following property:*

$$\frac{P(\{x\} \cup Z \succ y)}{P(\{y\} \cup Z \succ x)} = \frac{w_x}{w_y}. \tag{14}$$

It appears obvious that this axiom implies that this time the worst choice probabilities take the MNL form.

Theorem 3. *The reverse Luce's choice axiom (Axiom 2) holds iff the worst choice probabilities have the RMNL form given by:*

$$P(Y \succ x) = \frac{w_x}{w_x + w_Y}. \tag{15}$$

Proof. The proof goes along the same lines as the proof of Theorem 2. In a straightforward manner, we can systematically change the positions of the alternatives x and y, besides the subset Y in comparison to the preference $\succ$. $\square$

In the same way as outlined previously, we can combine our fundamental theorem, Theorem 1, with Theorem 3, which expresses the form of the worst choice probabilities, in order to obtain the form of the best choice probabilities.

Corollary 1. *Luce's axiom (Axiom 2) holds iff the best choice probabilities have the following form:*

$$P(x \succ Y) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} \frac{w_x}{w_x + w_{Y'}}.$$

Recall that this formulation of best choice probabilities was obtained by Anderson and de Palma [1999] who conceived it by inverting the signs of the random utility variables in an Additive Random Utility Model (ARUM). Our use here is rather in the spirit of fundamental theory with respect to more general Luce models and their ilk.

An impossibility result

It seems reasonable within the context of this paper to ask ourselves what model is produced when we simultaneously apply the two Axioms 1 and 2? We obtain the following:

Theorem 4 (Impossibility result)**.** *The probabilistic preferences satisfy simultaneously Axioms 1 and 2 iff:*

$$P(x \succ Y) = P(Y \succ x) = \frac{1}{1 + |Y|}. \tag{16}$$

Proof. If the conditions for Axioms 1 and 2 are met, the probabilities for binary choices is written as:

$$P(x \succ y) = \frac{b_x}{b_x + b_y} = \frac{w_y}{w_x + w_y},$$

this entails the equality of the following ratios $w_y/w_x = b_x/b_y$. Taking into account the constraints of normalization given by (5), we arrive at

$$w_x = \frac{1/b_x}{\sum_{y\in T} 1/b_y}. \tag{17}$$

We can now remind ourselves of the identity given by Eq. (8), which is applicable when we have three different alternatives x,y,$z \in T$. Therefore bringing us to:

$$1 - \frac{b_x}{b_x+b_y} - \frac{b_x}{b_x+b_z} + \frac{b_x}{b_x+b_y+b_z} - \frac{1/b_x}{1/b_x+1/b_y+1/b_z} = 0,$$

which can be rewritten as follows

$$1 - \frac{b_x}{b_x+b_y} - \frac{b_x}{b_x+b_z} + \frac{b_x}{b_x+b_y+b_z} - \frac{b_y b_z}{b_y b_z+b_x b_z+b_x b_y} = 0.$$

Regrouping the first three terms as well as the fourth and fifth terms, we obtain

$$\frac{b_y b_z - b_x^2}{(b_x+b_y)(b_x+b_z)} + \frac{b_x^2 b_z + b_x^2 b_y - b_y^2 b_z - b_y b_z^2}{(b_x+b_y+b_z)(b_y b_z+b_x b_z+b_x b_y)} = 0,$$

which can be further factorized as follows

$$(b_y b_z - b_x^2)\left(\frac{1}{(b_x+b_y)(b_x+b_z)} - \frac{b_y+b_z}{(b_x+b_y+b_z)(b_y b_z+b_x b_z+b_x b_y)}\right) = 0.$$

After some algebraic manipulations, the second term into parentheses can be factorized, obtaining

$$(b_y b_z - b_x^2)\frac{b_x b_y b_z}{(b_x+b_y)(b_x+b_z)(b_x+b_y+b_z)(b_y b_z+b_x b_z+b_x b_y)} = 0.$$

Therefore, we have necessary that: $b_x^2 = b_y b_z$. This type of equality is also valid for any permutation of the alternatives x,y,z, for instance providing us with $b_y^2 = b_x b_z$. Using the ratios of last two equality relationships, we can identify:

$b_x/b_y = 1$. In view of the normalization (5) we can deduct $b_x = 1/n$. Next, using Eq. (17), we have the same form $w_x = 1/n$. Finally, using the Eqs. (12) and (13), we obtain Eq. (16). $\square$

So this result gives us a strong impossibility result, describing how it is not possible to construct a model which satisfies Luce's two axioms at the same time - the original and the reverse - other than in a purely symmetric model, limiting this construction. We will discover in the next section that it is possible to obtain all the best and worse probabilities for each axiom thanks to a general identity that connects these probabilities.

4 Best&worst choice probabilities

We are now interested in the simultaneous choice of the best and worst alternatives. More formally, if x and z are two different alternatives of T, and Y is a non-empty subset of T that does not contain either of these two alternatives, we are interested in computing the probability that x is considered the best alternative while z is the worst alternative when the choice set contains x,Y, and z. The probabilities of these events will be called *best&worst choice probabilities* and are defined by

$$P(x \succ Y \succ z) \equiv P((x \succ Y) \cap (Y \succ z)). \tag{18}$$

The aim here is not to reconstruct an exhaustive theory of identities that can exist in all choice subsets of T as we have done for the best choice and the worst choice probabilities separately. Extensive discussions are made in Marley and Louviere [2005], in de Palma et al. [2017] and Delle Site et al. [2019].

It is also worth noting in the same vein of results and as a complement the interesting result of Colonius [2021] which gives necessary and sufficient conditions for a system of best&worst choice probabilities, i.e. all probabilities of the type $P(x \succ Y \succ z)$, so that they are rationalizable. The author thus obtains an extension of the result of Falmagne [1978] who showed that the non-negativity of the polynomials exhibited in Block and Marschak which are derived from the best choice pobabilities is a necessary and sufficient condition for rationalizability [see also the short proof of Fiorini, 2004]

Our goal here is rather to identify forms of best&worst choice probabilities that remain compatible with the Luce axiom for best choice probabilities (Axiom 1), or equivalently, models compatible with the reverse Luce axiom for worst choice probabilities (Axiom 2). These models could be interesting to test for future empirical applications.

Lemma 1. *Under Assumption 1, the best&worst choice probabilities are related to joint best choice probabilities by the following identity:*

$$P(x \succ Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} P((x \succ (Y \setminus Y') \cup \{z\}) \cap (z \succ Y')). \tag{19}$$

Similarly, the best&worst are related to joint worst choice probabilities by:

$$P(x \succ Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} P((Y' \succ x) \cap \{x\} \cup ((Y \setminus Y') \succ z)). \tag{20}$$

Proof. First, we note that by transitivity of the preference relation $\succ$, we have

$$(x \succ Y \succ z) = (x \succ Y \cup \{z\}) \cap (Y \succ z),$$

which implies, by using the indicator function $I(\cdot)$, the following equality

$$I(x \succ Y \succ z) = I(x \succ Y \cup \{z\}) \times I(Y \succ z).$$

Using the fact that

$$I(Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(z \succ Y'),$$

we get then

$$I(x \succ Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(x \succ Y \cup \{z\}) \times I(z \succ Y').$$

For any $y \in Y'$, if $z \succ y$ and $x \succ z$, we have necessarily $x \succ y$ which allows us to remove this redundancy by writing

$$I(x \succ Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(x \succ (Y \setminus Y') \cup \{z\}) \times I(z \succ Y').$$

Applying the mathematical expectation to both members of this equality and by linearity of expectation, we obtain Eq. (19).

Alternately, we can also write

$$I(x \succ Y \succ z) = I(x \succ Y) \cap (\{x\} \cup Y \succ z).$$

Then using the identity

$$I(x \succ Y) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(Y' \succ x),$$

and by similar arguments as above, we get

$$I(x \succ Y \succ z) = \sum_{\emptyset \subseteq Y' \subseteq Y} (-1)^{|Y'|} I(Y' \succ x) \times I(\{x\} \cup (Y \setminus Y') \succ z).$$

Applying the mathematical expectation then gives us Eq. (20). □

Thus, the best&worst choice probabilities in the formulation (19) appear as an alternating sum of joint best choice probabilities where z is preferred to a subset of Y while x is preferred to all the other alternatives, including z. In the formulation (20), it is rather an alternating sum of joint worst choice probabilities where x is worse than any subset of Y while z is worse than all the remaining alternatives, including x.

At this stage, Lemma 1 is not fully operational with such broad assumptions about preferences, even if we restrict ourselves to models where Axiom 1 (resp. Axiom 2) is satisfied. de Palma et al. [2017] have successfully used this type of identity to compute the best&worst choice probabilities in the generalized exteme value model.

The Gumbel-Strauss model

In this section we will compute the various choice probabilities seen above in a class of RUMs, i.e. in models where each alternative x is assigned a specific random utility U_x, where the best choice probabilities verify the axiom 1. We are interested in a particular RUM where the multivariate distribution of the utility vector is described by its CDF given by

$$P\left(\bigcap_{x\in T}(U_x \leq u_x)\right) = \exp\left[-\left(\sum_{x\in T} b_x e^{-\alpha u_x}\right)^{1/\alpha}\right], \tag{21}$$

where $\alpha \geq 1$. Thus, the margins are Gumbel distributions, and α allows to introduce correlations between utilities. We notice that the limiting case $\alpha = 1$ corresponds to the independence of the utilities.

We will refer to this model as the Gumbel-Strauss model insofar as the distribution of random utilities is a generalized Gumbel distribution, and that the authorship of the use of this distribution in the framework of RUMs is attributed to Strauss [1979].

We will again demonstrate that the best choice probabilities do indeed take the form MNL. The proof will be inspiring for upcoming results.

Lemma 2. *The best choice probabilities corresponding to the Gumbel-Strauss model, where the utilities have a CDF given by (21), take the MNL form given by (12).*

Proof. By definition, we have

$$P(x \succ Y) \equiv P\left(\bigcap_{y \in Y} (U_y \leq U_x)\right) = P(U_Y \leq U_x),$$

where $U_Y \equiv \max_{y \in Y} U_y$. The CDF of the couple composed of the variables U_x and U_Y and given by $F_2(u_x, u_Y) \equiv P(U_x \leq u_x, U_Y \leq u_Y)$, satisfies

$$F_2(u_x, u_Y) = \exp\left[-\left(b_x e^{-\alpha u_x} + b_Y e^{-\alpha u_Y}\right)^{1/\alpha}\right].$$

We then have to calculate the following double integral

$$P(x \succ Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{u_x} \frac{\partial^2 F_2(u_x, u_Y)}{\partial u_x \partial u_Y} du_Y du_x.$$

A first integration of the inner integral yields the following simplification

$$P(x \succ Y) = \int_{-\infty}^{+\infty} \frac{\partial F_2(u_x, u_x)}{\partial u_x} du_x.$$

As

$$\frac{\partial F_2(u_x, u_x)}{\partial u_x} = \frac{b_x}{b_x + b_Y} A e^{-u_x} \exp\left(-A e^{-u_x}\right).$$

where $A \equiv (b_x + b_Y)^{1/\alpha} > 0$, then using the fact that

$$\int_{-\infty}^{+\infty} Ae^{-u_x} \exp\left(-Ae^{-u_x}\right) du_x = 1,$$

we obtain

$$P(x \succ Y) = \frac{b_x}{b_x + b_Y}.$$

□

An immediate consequence of the previous result is the following one which gives the form of the worst choice probabilities for this model:

Proposition 2. *The worst choice probabilities corresponding to the Gumbel-Strauss model, take the alternate form given by (13).*

Proof. Since the best choice probabilities are of the MNL type, it is sufficient to use the Proposition 1 which provides the alternating form of the worst choice probabilities. □

We are now fully prepared for the computation of the simultaneous best and worst choice probabilities for the Gumbel-Strauss model. We have derived the following result:

Proposition 3. *The best&worst choice probabilities corresponding to the Gumbel-Strauss model, take the alternate form given by*

$$P(x \succ T \setminus \{x,z\} \succ z) = b_x \sum_{\emptyset \subseteq Z \subseteq T \setminus \{x,z\}} (-1)^{|Z|} \frac{b_z}{b_z + b_Z}. \tag{22}$$

Proof. We will use for our proof the Identity (19) of Lemma 1. It allows us to write the probabilities of the best and worst choice as follows

$$P(x \succ T \setminus \{x,z\} \succ z) = \sum_{\emptyset \subseteq Z \subseteq T \setminus \{x,z\}} (-1)^{|Z|} \theta_Z,$$

where

$$\theta_Z = \begin{cases} P(U_x \geq U_{T\setminus\{x,z\}}, & Z = \emptyset; \\ P(U_x \geq \max(U_X,U_z); U_z \geq U_Z), & \emptyset \subsetneq Z \subsetneq T \setminus \{x,z\}; \\ P(U_x \geq U_z \geq U_{T\setminus\{x,z\}}) & Z = T \setminus \{x,z\}, \end{cases}$$

where $X \equiv T \setminus \{x,z\} \setminus Z$. We have clearly

$$\theta_\emptyset = b_x \frac{b_z}{b_z + b_\emptyset}$$

since it corresponds to the probability that x is the best choice in T. We also have $\theta_{T\setminus\{x,z\}}$ is the probability that x is the best choice and z is the second-best choice in T. It satisfies

$$\theta_{T\setminus\{x,z\}} = P(U_z \geq U_{T\setminus\{x,z\}}) - P(U_z \geq U_{T\setminus\{z\}}) = \frac{b_z}{1-b_x} - b_z = b_x \frac{b_z}{b_z + b_{T\setminus\{x,z\}}}.$$

More generally, we shall prove that we always have $\theta_Z = b_x b_z/(b_z + b_Z)$. For $\emptyset \subsetneq Z \subsetneq T \setminus \{x,z\}$, we can then use the multivariate cumulative distribution function of the quadruplet (U_x,U_X,U_z,U_Z) which is given by

$$F_4(u_x,u_X,u_z,u_Z) =$$
$$\exp\big[-\big(b_x e^{-\alpha u_x} + b_X e^{-\alpha u_X} + b_z e^{-\alpha u_z} + b_Z e^{-\alpha u_Z}\big)^{1/\alpha}\big].$$

We need to compute the quadruple integral given below

$$\theta_Z = \int_{-\infty}^{+\infty} \int_{-\infty}^{u_x} \int_{-\infty}^{u_x} \int_{-\infty}^{u_z} \frac{\partial^4 F_4(u_x,u_X,u_z,u_Z)}{\partial u_x \partial u_X \partial u_z \partial u_Z} du_Z du_z du_X du_x.$$

By simplifying two inner integrals, we obtain a double integral

$$\theta_Z = \int_{-\infty}^{+\infty} \int_{-\infty}^{u_x} \frac{\partial^2 F_4(u_x,u_x,u_z,u_z)}{\partial u_x \partial u_z} du_z du_x.$$

By differentiating F_4 twice, we obtain

$$\frac{\partial^2 F_4(u_x,u_x,u_z,u_z)}{\partial u_x \partial u_z} = b_x b_z e^{-\alpha u_x} e^{-\alpha u_z} (A + \alpha - 1) A^{1-2\alpha} e^{-A},$$

where A is a function of the couple (u_x,u_z) defined by

$$A \equiv [(b_x + b_X)e^{-\alpha u_x} + (b_z + b_Z)e^{-\alpha u_z}]^{1/\alpha},$$

By performing a first change of variable at the level of the inner integral, and noticing that we have $b_x + b_X + b_z + b_Z = b_T = 1$, we get

$$\theta_Z = C \times b_x \frac{b_z}{b_z + b_Z}$$

where C is a constant equal to

$$C \equiv \int_{-\infty}^{\infty} \int_{e^{-u}}^{\infty} (A + \alpha - 1) A^{-\alpha} e^{-A} dA e^{-\alpha u} du.$$

We have thus proven that

$$P(x \succ T \setminus \{x,z\} \succ z) = C \times b_x \sum_{\emptyset \subseteq Z \subseteq T \setminus \{x,z\}} (-1)^{|Z|} \frac{b_z}{b_z + b_Z}.$$

It remains to show to finally prove our result that the constant C is indeed 1. Noting that the alternating sum is nothing but a worst choice probability, we have

$$P(x \succ T \setminus \{x,z\} \succ z) = C \times b_x P(T \setminus \{x,z\} \succ z).$$

By summing over all of the $z \in T \setminus \{x\}$ both members of the previous equation, we find

$$b_x = C \times b_x,$$

which indeed implies that $C = 1$. $\square$

As a conclusion to this section, we note that the best&worst probabilities for the Gumbel-Strauss model take the following form

$$P(x \succ T \setminus \{x,z\} \succ z) = P(x \succ T \setminus \{x\}) \times P(T \setminus \{x,z\} \succ z).$$

This multiplicative form specific to the logit model generated by independent Gumbel distributions, which corresponds to $\alpha = 1$ for the distribution given by (21), had been previously pointed out [Marley and Louviere, 2005, Proposition 9]. Here we generalize this interesting result in two ways: first, we determine the complete analytical form of the best&worst choice probabilities, and second, we show that this result remains valid even if the utilities are correlated in some way.

Alternatively, one can, without any further difficulty, consider the reverse model of the one given by Equation (22) which would have best&worst choice probabilities given by

$$P(x \succ T \setminus \{x,z\} \succ z) = w_z \sum_{\emptyset \subseteq X \subseteq T \setminus \{x,z\}} (-1)^{|X|} \frac{w_x}{w_x + w_X}. \tag{23}$$

5 Concluding remarks

In this paper we conducted a review of models that involve the best, the worst, and the best&worst choices in a choice set. We have focused on models that assume an underlying rationality, i.e. these probabilities are induced by a random order of preference over the alternatives.

It can be interesting in empirical applications to use models that have an explicit expression of these various probabilities. This is why we have concentrated our efforts on models of the MNL type where all the probabilities of interest to us are explicit and where a worst choice probability has an alternating form of best choice probabilities.

The best and worst probabilities considered simultaneously are the products of logit and alternating logit probabilities. We have also proposed reversed models where the worst probabilities have an MNL expression while the best probabilities have the alternating form. In future research, it would be interesting to implement these models in packages such as Apollo and to confront them with the current models already implemented, such as the MaxDiff model. This would open up interesting and new research paths.

References

Simon P. Anderson and André de Palma. Reverse discrete choice models. *Regional Science and Urban Economics*, 29(6):745–764, 1999.

Simon P. Anderson, André de Palma, and Jacques-François Thisse. *Discrete choice theory of product differentiation*. The MIT Press, Cambridge (Mass.), 1992.

Henry D. Block and Jacob Marschak. Random Orderings and Stochastic Theories of Response. Technical report.

Hans Colonius. A representation theorem for finite best–worst random utility models. *Journal of Mathematical Psychology*, 104:102596, 2021.

André de Palma, Karim Kilani, and Gilbert Laffond. Relations between best, worst, and best–worst choices for random utility models. *Journal of Mathematical Psychology*, 76:51–58, 2017.

Paolo Delle Site, Karim Kilani, Valerio Gatta, Edoardo Marcucci, and André de Palma. Estimation of consistent logit and probit models using best, worst and best–worst choices. *Transportation Research Part B: Methodological*, 128: 87–106, 2019.

Jean-Claude Falmagne. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18:52–72, 1978.

Samuel Fiorini. A short proof of a theorem of falmagne. *Journal of Mathematical Psychology*, 48(1):80–82, 2004.

Stephane Hess and David Palma. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32:100170, 2019. ISSN 1755-5345.

Jordan J. Louviere, Terry N. Flynn, and Anthony A. J. Marley. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press, Cambridge, 2015.

Robert Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York, 1959.

Anthony A. J. Marley. Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, 5(2):311–332, 1968.

Anthony A. J. Marley and Jordan J. Louviere. Some probabilistic models of best, worst, and best–worst choices. *Journal of Mathematical Psychology*, 49 (6):464–480, 2005.

Daniel McFadden. Econometric models of probabilistic choice. In C.F. Manski and D. McFadden, editors, *Structural analysis of discrete data with econometric applications*, pages 198–272. MIT Press, Cambridge, Mass., 1981.

Daniel McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.

David J. Strauss. Some results on random utility models. *Journal of Mathematical Psychology*, 20(1):35–52, 1979.