Best and worst choices

André de Palma, Karim Kilani, Gilbert Laffond

Best and worst choices

André de Palma[1] Karim Kilani[2] Gilbert Laffond[3]

May, 2013

[1]ENS Cachan, Centre d'Economie de la Sorbonne, Cachan 94230, France. Tel.: +33-6-63-64-4320; E-mail: andre.depalma@ens-cachan.fr

[2]CNAM, LIRSA, Paris 75003, France. Tel.: +33-1-40-27-2366; E-mail: karim.kilani@cnam.fr

[3]CNAM, LIRSA, Paris 75003, France. Tel.: +33-1-40-27-2366; E-mail: gilbert.laffond@cnam.fr

Abstract

We show that the number of individuals selecting their worst alternatives within a finite set of alternatives can be written as an alternating sum of the number of individuals having their best choice within subset of alternatives. The identities are then applied to random utility models, including the multinomial logit model, the mixed logit model and the disaggregated version of the CES representative consumer model. Finally, we show that better estimates are obtained if respondents are asked to reveal their worst instead of their best choices.
Keywords: Best-worst; CES; Discrete choice models; Gumbel distribution; Logit; Logsum
JEL classification: C25, D11

1 Introduction

There are several ways to express tastes in every day life. When asked to express preferences among alternatives, casual observation suggests that many respondents spontaneously tend to mention their best alternatives, and their most appreciated features (the most preferred hotel will be described with respect to the proximity to the sea, cleanliness or safety). Other respondents express their preferences by referring to their worst past experience (in the above example, they will enumerate the characteristics they dislike the most: spiders, dirtiness or noise). Similarly, young people often qualify the *cool* aspects of the current situation with respect to the opposite or worst situation, and will use negative expressions such as *no worries*. Exaggeration, maximization, minimization or understatements, very much used in English follow similar rhetoric.

Discrete choice models describe the probability that an alternative is chosen among a given set of alternatives (McFadden, 1974, 2001; Manski, 1977). Such models initially developed in Mathematical Psychology (Thurstone, 1927; Luce, 1959), Biophysics (Finney, 1947), Transportation (Warner, 1962; Ben-Akiva and Lerman, 1985), are now used in numerous fields, including Regional Science and Urban Economics (Anas and Chu, 1984), Labour Economics (Labeaga, Xisco and Spadaro, 2008), Marketing (Chintagunta and Nair, 2011) or Demography (Hoffman and Duncan, 1988). More recently, discrete choice models have played a key role in partial equilibrium models in Industrial Organization (see, Caplin and Nalebuff, 1991), as well as in General Equilibrium (see, Kokovin et al., 2012). These models rely on the assumption that preferences are expressed by best choice alternatives.

However, when an individual is asked to choose one among several products, the best and the worst choices provide equally useful benchmarks (Marley and Louviere, 2005; Marley, 2010). Sometimes, it is just not possible to ask for the best choice, but easier to ask for the worst choice (for sensitive questions in surveys, see Warner, 1965).

According to our main result, having information about preferences for the best choices over different subsets of alternatives, fully characterize preferences for the worst choice over the full set. More precisely, we derive an identity between best and worst choices. This result, valid for any set of probabilities, is based on an identity due to Henri Poincaré, the inclusion-exclusion principle in combinatorial mathematics. It justifies *de facto* the relevance of the worst choice in decision theory.

Besides the identity between best and worst choices, we explore the welfare properties of the worst choice, as a function of the welfare properties of the best choices. More generally, we explore how standard results in economics could be extended for the worst alternative. We show that under some circumstances, information on consumer surplus provides enough information to determine the welfare level of the worst possible choice, and we derive the welfare formula for the worst choices. Such welfare measures are proposed as a benchmark, and are shown to satisfy some version of Roy's identity. We also show how the Block and Marschak conditions (see Falmagne, 1978) can be extended to the worst alternative. Finally, we briefly compare the econometric efficiency of an estimator, when the respondent is asked to select their best choice, and their worst choice.

More formally, consider an individual facing a discrete number of alternatives. Probabilistic choice models derive expressions for the probability that a good is selected by a specific individual. Using our main theorem, we compute the probability that a specific alternative is the worst one. Discrete choice theory provides a parametrization of the choice probabilities of the best alternative. We focus on additive random utility models (ARUMs): $U_i = v_i + \varepsilon_i$, where U_i is the conditional random utility, with mean v_i and ε_i is an error term. Our main identity allows, without any extra hypothesis, to derive best choice probabilities (involving maximum operator on the random utilities) from the worst choice probabilities (involving minimum operator). We also consider the *reverse* additive random utility model $\widetilde{U}_i = v_i - \varepsilon_i$, which are shown to be closely related to worst choice probabilities. Similarly, the benefit of the best alternative is logically related to the benefit of the worst ones.

In Section 2, we show that the number of agents choosing their best choice within a set of alternatives can be written as an alternating sum of the number of agents choosing their worst choice over appropriately chosen subsets (Theorem 1). Section 3 provides the probabilistic version of Theorem 1 and extends it to the joint best and worst choice probabilities. Section 4 considers the ARUM specification and provides the relationship between the stochastic maximum and minimum utilities. Section 5 considers the standard multinomial logit model as an example of ARUM. Section 6 considers several extensions of the multinomial logit: the reversed logit, the GEV, the mixed logit, the CES. In particular, In particular, Falmagne's characterization is extended to the worst case. The relation between the best choice, the worst choice and the reverse choice is provided. Section 7 provides numerical examples and compare situations where hypothetical respondents are asked to reveal their best or their worst choices.

2 Worst versus best choices

Consider N agents facing a finite choice set $\mathcal{T}$ of n alternatives, $n \geq 2$. Their preferences are described by a linear order $\succeq$ on $\mathcal{T}$. The corresponding ranking of the alternatives from the best to the worst choice is denoted by ρ. Let Ω be the set of all rankings ρ; it has $n!$ elements. Let $B_{\mathcal{S}}^{\#}(i)$ be the number of agents with i as their best choice in $\mathcal{S} \subset \mathcal{T}$ with $\sum_{i \in \mathcal{T}} B_{\mathcal{S}}^{\#}(i) = N$, while $W_{\mathcal{S}}^{\#}(i)$ denotes the number of agents with i as their worst choice in $\mathcal{S}$. In Theorem 1, we derive an identity between $W_{\mathcal{T}}^{\#}(i)$ and $B_{\mathcal{S}}^{\#}(i)$, $\mathcal{S} \subset \mathcal{T}$.

2.1 Preliminary example

Consider an election with three candidates i, j, k and $N = 100$ voters. There are 6 rankings of the candidates with the following ballots:

Rankings	ijk	ikj	jik	jki	kij	kji
Ballots	12	30	10	28	2	18

The ranking in which $i \succ j \succ k$ is denoted by ijk, etc., where $\succ$ denotes the strict linear order.

If each voter is asked to select his favorite candidate (plurality or relative majority), candidate i polled 42 votes, j polled 38 votes, and k polled 20 votes. Candidate i is elected. When considering a vote by elimination, k polled 22 votes, j polled 32 votes, and i polled 46 votes. Candidate k is elected.

This simple example shows that the knowledge of the number of voters for the best candidate cannot be used to identify the most disproved candidate. The additional information required is the number of voters in pairwise comparisons: $B^{\#}_{\{i,j\}}(i) = 44$, $B^{\#}_{\{i,k\}}(i) = 52$, since $W^{\#}_{\{i,j,k\}}(i) = N - B^{\#}_{\{i,j\}}(i) - B^{\#}_{\{i,k\}}(i) + B^{\#}_{\{i,j,k\}}(i)$. Indeed, in the sum $W^{\#}_{\{i,j,k\}}(i) + B^{\#}_{\{i,j\}}(i) + B^{\#}_{\{i,k\}}(i)$, all the agents are involved at least once. Agents with rankings ijk and ikj are appearing twice. Their number corresponds to $B^{\#}_{\{i,j,k\}}(i)$, the trinary best choice of i in $\{i, j, k\}$, implying the former identity.

We wish to find an identity which expresses the number agents selecting i as the best alternative as a function of the number of agents selecting i as their worst alternatives in appropriately chosen choice sets. The analogous identity relates, in a similar manner, the number of agents selecting i as their worst alternatives as a function of number of agents selecting alternatives i as their best alternative in appropriately chosen choice sets.

2.2 Best worst identity

The reasoning in Section 2.1 can be extended to any number of alternatives. The proof relies on an inclusion-exclusion principle in probability theory attributed to Poincaré (see Poincaré, 1912, pp. 58-60).[1] We have:

Theorem 1 *The number $W_{\mathcal{T}}^{\#}(i)$ of agents choosing i as their worst alternative in $\mathcal{T}$ can be expressed as an alternate sum of the numbers $B_{\mathcal{S}}^{\#}(i)$ of agents choosing i as their best alternative in $\mathcal{S}$:*

$$W_{\mathcal{T}}^{\#}(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} B_{\mathcal{S}}^{\#}(i), \; i \in \mathcal{T}. \tag{1}$$

Proof. For a given agent with ranking $\rho \in \Omega$, let $I_\rho(\cdot)$ the operator verifying: $I_\rho(E) = 1$ if $\rho \in E$, and $I_\rho(E) = 0$ otherwise, $E \subset \Omega$. Let $\chi_{\mathcal{S}}^{B}(i) \equiv I_\rho\left(\bigcap_{k\in\mathcal{S}} E_{ik}\right)$ be the best choice indicators of i in $\mathcal{S}$, where $E_{ik} \equiv \{\rho \in \Omega : i \succeq k\}$. In the same way, let $\chi_{\mathcal{S}}^{W}(i) \equiv I_\rho\left(\bigcap_{k\in\mathcal{S}} E_{ki}\right)$ be the worst choice indicators of i in $\mathcal{S}$. Notice that $I_\rho(\cdot)$ has properties similar to a probability operator $P(\cdot)$ defined on the power set of Ω. Indeed, $I_\rho(\Omega) = 1$, and $I_\rho(E_1 \cup E_2) = I_\rho(E_1) + I_\rho(E_2)$, $E_1, E_2 \subset \Omega$ with $E_1 \cap E_2 = \emptyset$. Therefore, using the inclusion-exclusion principle

$$P\left(\bigcap_{k\in\mathcal{T}} E_{ki}\right) = 1 - \sum_{\mathcal{R}\subset\mathcal{T}} (-1)^{|\mathcal{R}|-1} P\left(\bigcap_{k\in\mathcal{R}} E_{ki}^{C}\right),$$

[1] The inclusion-exlusion principle dates back to 1718 and is attributed to de Moivre. It is also related to Bernoulli numbers published in *Ars Conjectandi*, 1713. The standard inclusion-exlusion principle has been extended by Möbius in 1832 in the field of number theory (see the presentation and examples of Möbius inversion in Bender and Goldman, 1975). Fiorini (2004) used Möbius inversion to provide a simpler proof of the Falmagne's theorem. The relation between best and worst choices and the Block-Marschak polynomials used in Falmagne's theorem is discussed in Section 6.3.

where E^C denotes the complementary of $E \subset \Omega$. Thus,

$$I_\rho\left(\bigcap_{k\in\mathcal{T}} E_{ki}\right) = 1 - \sum_{\mathcal{R}\subset\mathcal{T}} (-1)^{|\mathcal{R}|-1} I_\rho\left(\bigcap_{k\in\mathcal{R}} E_{ki}^C\right).$$

Notice that $I_\rho\left(\bigcap_{k\in\mathcal{T}} E_{ki}\right) = \chi_{\mathcal{T}}^W(i)$. Moreover, since $E_{ki}^C = E_{ik}$, $\forall k \neq i$, and using the fact that $E_{ii} = \{\rho \in \Omega : i \succeq i\} = \Omega$, so that $E_{ii}^C = \emptyset$, we get

$$\chi_{\mathcal{T}}^W(i) = 1 - \sum_{\mathcal{R}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{R}|-1} I_\rho\left(\bigcap_{k\in\mathcal{R}} E_{ik}\right).$$

The above equation can be rewritten as

$$\chi_{\mathcal{T}}^W(i) = \sum_{\emptyset\subseteq\mathcal{R}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{R}|} I_\rho\left(\bigcap_{k\in\mathcal{R}\cup\{i\}} E_{ik}\right).$$

Posing $\mathcal{S} = \mathcal{R} \cup \{i\}$, we get

$$\chi_{\mathcal{T}}^W(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} I_\rho\left(\bigcap_{k\in\mathcal{S}} E_{ik}\right).$$

Since $I_\rho\left(\bigcap_{k\in\mathcal{S}} E_{ik}\right) = \chi_{\mathcal{S}}^B(i)$, for $\{i\} \subset \mathcal{S} \subset \mathcal{T}$, we obtain

$$\chi_{\mathcal{T}}^W(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \chi_{\mathcal{S}}^B(i), \; i \in \mathcal{T}. \tag{2}$$

Identity (1) is obtained by summing-up both sides of (2) over the N agents and by permutation of the sum signs of its RHS.

The functional form (1) involves a sum with terms alternating in signs. For example, with four alternatives, we obtain $W_{\{i,j,k,l\}}^{\#}(i) = N - B_{\{i,j\}}^{\#}(i) -$

$B^{\#}_{\{i,k\}}(i) - B^{\#}_{\{i,l\}}(i) + B^{\#}_{\{i,j,k\}}(i) + B^{\#}_{\{i,j,l\}}(i) + B^{\#}_{\{i,k,l\}}(i) - B^{\#}_{\{i,j,k,l\}}(i)$.

Symmetrically, inverting the side of the preference order $\succeq$ into $\preceq$, one obtains the opposite identity relating the best choice numbers $B^{\#}_{\mathcal{T}}(i)$, to the worst choice ones $W^{\#}_{\mathcal{S}}(i)$ as follows

$$B^{\#}_{\mathcal{T}}(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} W^{\#}_{\mathcal{S}}(i), \; i \in \mathcal{T}. \tag{3}$$

We now consider the case of a large number of agents.

3 Large number of agents

3.1 Worst choice probabilities

So far, we have considered a finite number N of agents. The agent's preferences are described by a ranking ρ over the n alternatives. Below, we analyze the limit case $N \to \infty$, and study the asymptotic behavior of the fraction of agents selecting i as their best or worst choice in $\mathcal{S} \subset \mathcal{T}$.

We assume that there are variables non-observable by the modeler which determine individual choices. As a consequence, agent preferences can only be described up to a probability distribution. Let $P(\rho)$ denote the probability that the agent ranking of the alternatives is ρ. Following Block and Marschak (1960), the best choice probability $B_{\mathcal{S}}(i)$ of i in $\mathcal{S}$ and the ranking probabilities are related as follows: $B_{\mathcal{S}}(i) = \sum_{\rho\in\Omega:\ i\succeq k,\ \forall k\in\mathcal{S}} P(\rho)$. We define $W_{\mathcal{S}}(i)$ in a similar manner, $i \in \mathcal{S} \subset \mathcal{T}$.

Assume the agents are statistically independent and identical. In this case, the number of agents $B^{\#}_{\mathcal{S}}(i)$, has a binomial distribution. According to the law of large numbers:

$$\lim_{N\to\infty} \frac{B^{\#}_{\mathcal{S}}(i)}{N} = B_{\mathcal{S}}(i), \text{ a.e., } i \in \mathcal{S} \subset \mathcal{T}. \tag{4}$$

Thus, $B_{\mathcal{S}}(i)$ denotes the probability that an agent, randomly selected in the population selects alternative $i \in \mathcal{S}$ as his/her best alternative in $\mathcal{S}$, while $W_{\mathcal{S}}(i)$ denotes the probability that an agent selects alternative i as his/her worst alternative in $\mathcal{S}$. Preferences over alternatives are described by a probability $P(\cdot)$. The identity for the best and worst choice probabilities is a straightforward application of Theorem 1:

Corollary 1 *The worst choice probability of alternative i in $\mathcal{T}$ can be expressed as an alternate sum of best choice probabilities:*

$$W_{\mathcal{T}}(i) = \sum_{\{i\} \subset \mathcal{S} \subset \mathcal{T}} (-1)^{|\mathcal{S}|-1} B_{\mathcal{S}}(i), \; i \in \mathcal{T}. \tag{5}$$

Proof. See Appendix B.

Corollary 1 states that the system of best choice probabilities,[2] $B_{\mathcal{S}}(i)$, $i \in \mathcal{S} \subset \mathcal{T}$, and the system of worst choice probabilities $W_{\mathcal{S}}(i)$, $i \in \mathcal{S} \subset \mathcal{T}$, are isomorphic. Preferences with respect to the worst choice can be elicited by considering only preferences with respect to the best choice within different appropriately chosen choice subsets.

The counterpart of (3) is

$$B_{\mathcal{T}}(i) = \sum_{\{i\} \subset \mathcal{S} \subset \mathcal{T}} (-1)^{|\mathcal{S}|-1} W_{\mathcal{S}}(i), \; i \in \mathcal{T}. \tag{6}$$

This identity will be used in Section 6.2 where we consider reverse random utility models.

[2] The system has $\sum_{r=2}^{n} \binom{n}{r} (r-1) = n2^{n-1} - 2^n + 1$ degrees of freedom.

3.2 Best-worst choice probabilities

During the last years, there was a growing interest in best–worst scaling experiments, where a person is asked to select both the best and the worst alternative in a set of alternatives (see e.g. Marley and Louviere, 2005). In this vein, let us denote by $BW_{\mathcal{T}}(i,j)$ the joint best-worst probability, that i is the best and j the worst alternative in $\mathcal{T}$:

$$BW_{\mathcal{T}}(i,j) \equiv P\left(\bigcap_{k\in\mathcal{T}} \{\rho : i \succeq k \succeq j\}\right),\ i,j \in \mathcal{T},\ i \neq j.$$

Notice that in the trinary case, the best-worst choice probabilities can be obtained from the best choice probabilities (binary and trinary). Indeed, we have $BW_{\{i,j,k\}}(i,j) = P(ikj)$. Using the fact that: $B_{\{j,k\}}(j) = P(jik) + P(jki) + P(ijk)$, and that: $B_{\{i,j,k\}}(j) = P(jik) + P(jki)$, we obtain $BW_{\{i,j,k\}}(i,j) = B_{\{j,k\}}(j) - B_{\{i,j,k\}}(j)$. However, the expressions are more involved for $n \geq 4$. General expressions for the best-worst choice probabilities are given in:

Theorem 2 *The Best-worst choice probabilities $BW_{\mathcal{T}}(i,j)$, $i,j \in \mathcal{T}$, $i \neq j$, and best and wort choice probabilities satisfy:*

$$(7) \qquad BW_{\mathcal{T}}(i,j) = B_{\mathcal{T}}(i)\, W_{\mathcal{T}-\{i\}}(j) + \sum_{\{j\}\subset\mathcal{S}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{S}|-1} c_{\mathcal{S}}(i,j),$$

where $W_{\mathcal{T}-\{i\}}(j)$ are given by (5), and where $c_{\mathcal{S}}(i,j) \equiv cov\left(\chi^B_{\mathcal{T}}(i), \chi^B_{\mathcal{S}}(j)\right)$, with $\chi^B_{\mathcal{S}}(i) = 1$ if $i \succeq k$, $\forall k \in \mathcal{S}$, and $\chi^B_{\mathcal{S}}(i) = 0$ otherwise.

Proof. Define the joint best-worst indicator such that i is the best and j the worst choice in $\mathcal{T}$ as

$$\chi^{BW}_{\mathcal{T}}(i,j) \equiv \begin{cases} 1, & \text{if } i \succeq k \succeq j,\ k \in \mathcal{T}; \\ 0, & \text{otherwise, } i,j \subset \mathcal{T},\ i \neq j. \end{cases}$$

It is related to the best and the worst indicators since

$$\chi_{\mathcal{T}}^{BW}(i,j) = \chi_{\mathcal{T}}^{B}(i)\,\chi_{\mathcal{T}-\{i\}}^{W}(j)\,,\; i,j \in \mathcal{T},\, i \neq j. \tag{8}$$

Indeed, using the operator $I_\rho(\cdot)$ (see above), we have

$$\chi_{\mathcal{T}}^{BW}(i,j) = I_\rho\left(\bigcap_{k\in\mathcal{T}}\{i \succeq k\} \bigcap_{k\in\mathcal{T}-\{i\}}\{k \succeq j\}\right),$$

Using the fact that $I_\rho(E \cap F) = I_\rho(E) \times I_\rho(F)$, $E, F \subset \Omega$, Eq. (8), is obtained. Applying the expectation operator on both sides of (8), we get

$$BW_{\mathcal{T}}(i,j) = \mathbb{E}\left[\chi_{\mathcal{T}}^{BW}(i,j)\right] = \mathbb{E}\left[\chi_{\mathcal{T}}^{B}(i)\,\chi_{\mathcal{T}-\{i\}}^{W}(j)\right],$$

which can be rewritten as

$$BW_{\mathcal{T}}(i,j) = B_{\mathcal{T}}(i)\,W_{\mathcal{T}-\{i\}}(j) + cov\left(\chi_{\mathcal{T}}^{B}(i)\,,\chi_{\mathcal{T}-\{i\}}^{W}(j)\right).$$

Applying (2), we have $\chi_{\mathcal{T}-\{i\}}^{W}(j) = \sum_{\{j\}\subset\mathcal{S}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{S}|-1}\,\chi_{\mathcal{S}}^{B}(j)$. Therefore, using (8), the best-worst choice probabilities are

$$BW_{\mathcal{T}}(i,j) = B_{\mathcal{T}}(i)\,W_{\mathcal{T}-\{i\}}(j) + cov\left(\chi_{\mathcal{T}}^{B}(i)\,, \sum_{\{j\}\subset\mathcal{S}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{S}|-1}\,\chi_{\mathcal{S}}^{B}(j)\right),$$

which can be rewritten, using the linearity of the covariance operator, as in Eq. (7).

Note that if $\chi_{\mathcal{T}}^{B}(i)$ and $\chi_{\mathcal{S}}^{B}(j)$, $\mathcal{S} \subset \mathcal{T} - \{i\}$, are two independent random variables, the formula reduces to $BW_{\mathcal{T}}(i,j) = B_{\mathcal{T}}(i)\,B_{\mathcal{T}-\{i\}}(j)$, an independence result. In the logit framework, independency holds allowing explicit expressions for $BW_{\mathcal{T}}(i,j)$. In general, the best choice indicators $\chi^B_{\mathcal{T}}(i)$ and $\chi^B_{\mathcal{S}}(j)$ are correlated, implying that the best-worst choice probabilities cannot be deduced from the system of best choice probabilities. Computation of $(n+1)(n-2)/2$ covariance terms $c_{\mathcal{S}}(i,j)$, with $c_{\mathcal{S}}(i,j) = \mathbb{E}\left[\chi^B_{\mathcal{T}}(i)\,\chi^B_{\mathcal{S}}(j)\right] - B_{\mathcal{T}}(i)\,B_{\mathcal{S}}(j)$, is involved. The terms $\mathbb{E}\left[\chi^B_{\mathcal{T}}(i)\,\chi^B_{\mathcal{S}}(j)\right]$ can be explicitly performed when more structure is provided for the choice probabilities as it is the case in the ARUM framework addressed in the next section.

4 Additive random utility models

In random utility models or ARUMs (see e.g. McFadden, 1974 and 2001), it is assumed that each alternative k has a conditional utility:[3] $U_k = v_k + \epsilon_k$, where v_k is a systematic part and ϵ_k is a random error term, $k \in \mathcal{T}$. The distribution of the vector of random error terms $\boldsymbol{\varepsilon}^{\mathcal{T}}$ with components ε_k, $k \in \mathcal{T}$, is assumed to be absolutely continuous with respect to the Lebesgue measure over R^n. We denote by $F^{\mathcal{T}}(\cdot)$ the CDF of $\boldsymbol{\varepsilon}^{\mathcal{T}}$. This hypothesis is standard in ARUMs and will be used in several proofs below.

The preference of any specific agent is described by the realization of random variables U_k, $k \in \mathcal{T}$. For example, individual l faces realizations U^l_k, and prefers i over k $(i \succeq k)$ in $\mathcal{S}$ iff $U^l_i \geq U^l_k$. As a consequence, agent l selects i as his/her best choice in $\mathcal{S}$ iff $U^l_i \geq U^l_k$, $\forall k \in \mathcal{S}$. In a probabilistic setting, choices are described by the law of comparative judgments (Thurstone, 1929). An agent randomly selected in the population (agents are assumed to be statistically independent) chooses alternative i with a probability

$$B_{\mathcal{S}}(i) = \Pr\left(v_i + \varepsilon_i \geq v_k + \varepsilon_k,\ \forall k \in \mathcal{S}\right),\ i \in \mathcal{S}. \tag{9}$$

[3] The terminology "conditional utility" is somewhat an abuse of language, but it is commonly used in the discrete choice litterature. The rationalization of preferences in discrete choice models has been discussed by Small and Rosen (1981).

Note that since $\boldsymbol{\varepsilon}^{\mathcal{T}}$ is assumed absolutely continuous, ties occur with zero probability so that the above expression is non ambiguous. Likewise, the worst choice probabilities are given by

$$W_{\mathcal{S}}(i) = \Pr\left(v_i + \varepsilon_i \leq v_k + \varepsilon_k,\ \forall k \in \mathcal{S}\right),\ i \in \mathcal{S}. \tag{10}$$

4.1 Choice probabilities

In order to recover the worst choice probabilities, Corollary 1 can be applied to any system of best choice probabilities $B_{\mathcal{S}}(i)$, $i \in \mathcal{S} \subset \mathcal{T}$, derived for ARUMs such as the logit (see Section 5), the probit, the generalized extreme value , or the mixed logit models.[4] This means that any ARUM which provides the best choice probability can be associated to the worst choice probability with no additional hypothesis and simple computations (involving no additional integrals).

In general, numerical values of the probabilities $B_{\mathcal{S}}(i)$, $i \in \mathcal{S}$, amount to compute $|\mathcal{S}| - 1$ integrals, which is rather intricate for $|\mathcal{S}| \geq 4$. In some specific cases, these probabilities have a closed form. The most standard case is the logit considered in Section 5.

For ARUMs, standard properties of the partial derivatives of the best choice probabilities (see Anderson, de Palma, and Thisse, 1992) can be easily extended to worst choice probabilities. The signs of the best choice probabilities derivatives with respect to the systematic parts are: $\partial B_{\mathcal{S}}(i)/\partial v_i > 0$; $\partial B_{\mathcal{S}}(i)/\partial v_k < 0$, $k \neq i$. It can be shown that signs are reversed for the worst choice probabilities derivatives: $\partial W_{\mathcal{S}}(i)/\partial v_i < 0$; $\partial W_{\mathcal{S}}(i)/\partial v_k > 0$, $k \neq i$.[5]

[4] For mixed logit models, the systematic parts of the utilities are themselves randomly distributed.

[5] For ARUM's, the symmetry of the partial derivatives of the best choice probabilities states that: $\partial B_{\mathcal{S}}(i)/\partial v_k = \partial B_{\mathcal{S}}(k)/\partial v_i$. A similar property holds for the worst choice probabilities: $\partial W_{\mathcal{S}}(i)/\partial v_k = \partial W_{\mathcal{S}}(k)/\partial v_i$.

4.2 Minimum-maximums utilities

The expected maximum utility $\mathbb{E}\left[\max_{k\in\mathcal{T}}(v_k+\varepsilon_k)\right]$ is often used in discrete choice models as an approximate welfare measure (see Small and Rosen, 1981). The formula is exact in the absence of income effects. Otherwise, the welfare measures are more involved (see e.g. de Palma and Kilani, 2011). We consider here an alternative measure, the expected minimum utility $\mathbb{E}\left[\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)\right]$ which corresponds to the benefit of the worst situation. Facing uncertainty, the worst case scenario provides useful information when resorting, for example, to minimum regret theory. Herbert Simon introduced the theory of Minimal Regret. Accordingly, the agent minimizes an expression based on his/her regret.

Consider the random variables $\max_{k\in\mathcal{S}}(v_k+\varepsilon_k)$ and $\min_{k\in\mathcal{S}}(v_k+\varepsilon_k)$, the best and worst utility in $\mathcal{S}\subset\mathcal{T}$, respectively. They satisfy the following identity (recall error terms are absolutely continuous with respect to the Lebesgue measure over R^n):

Theorem 3 *For an ARUM, the worst and best utilities satisfy:*

$$\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)=\sum_{\mathcal{S}\subset\mathcal{T}}(-1)^{|\mathcal{S}|-1}\max_{k\in\mathcal{S}}(v_k+\varepsilon_k). \tag{11}$$

Proof. The minimum utility can be rewritten using the worst choice indicators

$$\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)=\sum_{k\in\mathcal{T}}(v_k+\varepsilon_k)\,W_{\mathcal{T}}(k).$$

Using identity (2), we get

$$\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)=\sum_{k\in\mathcal{T}}(v_k+\varepsilon_k)\sum_{\{k\}\subset\mathcal{S}\subset\mathcal{T}}(-1)^{|\mathcal{S}|-1}B_{\mathcal{S}}(k).$$

Inverting the two sum signs we get

$$\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\sum_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)B_{\mathcal{S}}\left(k\right).$$

Since: $\sum_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)B_{\mathcal{S}}\left(k\right)=\max_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)$, we get Eq. (11).

When utilities have finite expectation, identity (11) implies

$$\mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\mathbb{E}\left[\max_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)\right]. \tag{12}$$

For ARUMs, Roy's identity holds and is written as (see McFadden, 1981)

$$\frac{\partial\mathbb{E}\left[\max_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)\right]}{\partial v_i}=B_{\mathcal{S}}\left(i\right),\ \ i\in\mathcal{S}\subset\mathcal{T}. \tag{13}$$

We show below an identity analog to Roy's identity which holds for the expected minimum utility. We have:

Corollary 2 *For an ARUM where utilities have finite expectation, we have:*

$$\frac{\partial\mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]}{\partial v_i}=W_{\mathcal{T}}\left(i\right),\ i\in\mathcal{T}. \tag{14}$$

Proof. See Appendix B.

Accordingly, the quantity $\mathbb{E}\left[\max_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]-\mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]$ can be interpreted as a variability measure for a decision maker facing the choice set $\mathcal{T}$. This index is worth considering in nested models, where the consumers consider alternative choice sets (see e.g. Ben-Akiva and Lerman, 1985).

5 The multinomial logit

In this section, we illustrate the previous results in the multinomial logit (MNL) model framework. In this case, we show that choice probabilities (best and worst) as well as expected (maximum/minimum) utility have explicit functional forms.

5.1 Best choice logit

The standard multinomial logit model assumes that the error terms ε_k are i.i.d. Gumbel so with CDF given by: $F(x) = \exp(-\exp(\gamma - x))$, where γ is Euler's constant introduced for normalization purpose. The choice probabilities have closed form known as the MNL formula:

$$B_{\mathcal{S}}(i) = \frac{e^{v_i}}{\sum_{k \in \mathcal{S}} e^{v_k}}, \; i \in \mathcal{S} \subset \mathcal{T}. \tag{15}$$

Note that $B_{\mathcal{S}}(i)$ given by (15) are nondecreasing in v_i and nonincreasing in v_k, $i, k \in \mathcal{S}$, $k \neq i$, and is log-concave[6] (see Caplin and Nalebuff, 1991). The log-concavity of the best choice probabilities are essential for studying equilibrium existence in an imperfect competition framework (see Anderson, de Palma and Thisse, 1992). Below, we show that these properties can be extended to worst choices.

5.2 Worst choice logit

Corollary 1 derives a new closed form formula for the worst choice probabilities. In the case of the logit, we have an explicit formula for the probability that i is the worst choice. It is given by:

[6] This means that $\ln B_{\mathcal{S}}(i)$ is concave in the vecor of systematic utilities with components v_k, $k \in \mathcal{S}$.

Proposition 1 *For the multinomial logit model, the worst choice probabilities are given by the worst multinomial logit (WMNL) formula:*

$$W_{\mathcal{T}}(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \frac{e^{v_i}}{\sum_{k\in\mathcal{S}} e^{v_k}},\ i\in\mathcal{T}. \tag{16}$$

Moreover, $W_{\mathcal{T}}(i)$ *is log-concave in the vector of systematic utilities* v_k, $k\in\mathcal{T}$.

Proof. See Appendix A.

Explicit expressions for the worst choice probabilities and log-concavity are important for estimation when preferences are elicited using worst choice experiments, as discussed in Section 7.

5.3 Best-worst choice logit

The elicitation of preferences can also be made using questions on the best and worst choices. Several empirical applications have been made in this direction by Marley and Louviere (2005) and Marley (2010).

Lemma 1 *For the MNL, we have* $c_{\mathcal{S}}(i,j)=0$, $j\in\mathcal{S}\subset\mathcal{T}-\{i\}$, $i\in\mathcal{T}$.

Proof. See Appendix B.

We can now derive a closed form for the best-worst choice probabilities in the case of the MNL. Using Corollary 2 and Lemma 1, we obtain $BW_T(i,j) = B_{\mathcal{T}}(i)\,W_{\mathcal{T}-\{i\}}(j)$. Therefore, from the MNL formula (15) and the WMNL formula (16), we get:

Proposition 2 *For the multinomial logit model, the joint best-worst choice probabilities are given by the best-worst multinomial logit (BWMNL) formula:*

$$BW_T(i,j) = \frac{e^{v_i}}{\sum\limits_{k\in\mathcal{T}} e^{v_k}} \sum_{\{j\}\subset\mathcal{S}\subset\mathcal{T}-\{i\}} (-1)^{|\mathcal{S}|-1} \frac{e^{v_j}}{\sum\limits_{k\in\mathcal{S}} e^{v_k}},\ i,j\in\mathcal{T},\ i\neq j. \tag{17}$$

Proof. See Appendix A.

Formula (17) is a consequence of the fact that the random variables $\chi_{\mathcal{T}}^{B}(i)$ and $\chi_{\mathcal{T}-\{i\}}^{W}(j)$ are independent. This independence property was already noticed by Marley and Louviere (2005, Proposition 9). They show that independence holds for ARUMs with i.i.d. random error terms iff the error terms are Gumbel distributed (as for the logit). Independence of error terms do not imply in general independency. The exact explicit formula (17) is new in the discrete choice models literature. Notice that since the best-worst choice probability $BW_{\mathcal{T}}(i,j)$ is a product of a best choice probability $B_{\mathcal{T}}(i)$ and worst choice probability $W_{\mathcal{T}-\{i\}}(j)$ which are both log-concaves, $BW_{\mathcal{T}}(i,j)$ is also log-concave.

5.4 Best and worst logsums

In the MNL, the expected maximum utility is given by the standard logsum formula

$$\mathbb{E}\left[\max_{k\in\mathcal{S}}(v_k+\varepsilon_k)\right]=\ln\sum_{k\in\mathcal{S}}e^{v_k},\ \mathcal{S}\subset\mathcal{T}. \tag{18}$$

This formula has been used not only in economics but also in Transportation, Urban Economics, or Geography. We introduce below an analogous (yet more complex) explicit formula for the expected minimum utility. We refer to it as the *worst logsum* formula:

Proposition 3 *For the MNL, the expected minimum utility is given by the worst logsum formula:*

$$\mathbb{E}\left[\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)\right]=\sum_{\mathcal{S}\subset\mathcal{T}}(-1)^{|\mathcal{S}|-1}\ln\sum_{k\in\mathcal{S}}e^{v_k}. \tag{19}$$

Proof. See Appendix A.

In the symmetric case where $v_k = v$, $k \in \mathcal{T}$, we get $\mathbb{E}\left[\max_{k\in\mathcal{S}}(v+\varepsilon_k)\right] = v + \ln q$ (see Eq. 18), where $q \equiv |\mathcal{S}|$, which is nondecreasing and concave in q. A routine computation shows that

$$\mathbb{E}\left[\min_{k\in\mathcal{T}}(v+\varepsilon_k)\right] = v + \textstyle\sum_{q=1}^{n}(-1)^{q-1}\binom{n}{q}\ln q,$$

which is strictly decreasing and convex in n.

6 Extensions

6.1 GEV models

McFadden has extended the MNL model to the generalized extreme value (GEV) model where the error terms ε_k, $k \in \mathcal{T}$, are Gumbel distributed but allow for correlation via GEV generator functions. The best choice probabilities for GEV models have a closed form: $B_{\mathcal{S}}(i) = \partial \ln G_{\mathcal{S}} / \partial e^{v_i}$, $i \in \mathcal{S} \subset \mathcal{T}$, where $G_{\mathcal{S}}$ are GEV generator functions which are function of e^{v_k}, $k \in \mathcal{S}$. Applying Corollary 1, the worst choice probabilities for GEV are:

$$W_{\mathcal{T}}(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \frac{\partial \ln G_{\mathcal{S}}}{\partial e^{v_i}}, \ i \in \mathcal{T}. \tag{20}$$

No closed form for the best-worst choice probabilities (except for MNL models which are special cases of GEV models) exists. However, they can be approached in two ways. First, by sampling the error terms ε_k, $k \in \mathcal{T}$, using Metropolis-Hastings sampler (see Train, 2003). Alternatively, they can rely on the calculation of the covariances $c_{\mathcal{T}}(i,j)$ involved in formula (7) of Corollary 2, performing numerical calculations of double-integrals.

Recall that the expected maximum utility in GEV models is explicit and provides an extension of the logsum formula (18): $\mathbb{E}\left[\max_{k\in\mathcal{S}}(v_k+\varepsilon_k)\right] = \ln G_{\mathcal{S}}$, $\mathcal{S} \subset \mathcal{T}$. Therefore, application of (12) yields a generalization of the worst logsum formula (19): $\mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\ln G_{\mathcal{S}}$. An analogous formula was derived analytically by Misra (2005) for the reverse GEV model. The relation between reverse model and minimizing behavior is analyzed in the next section.

6.2 Reverse models

Anderson and de Palma (1999) have broaden the class of standard ARUMs to their reverse counterpart, referred to as *reverse discrete choice models*. The random error terms of the utility are simply shifted so that the utility in the reverse model is: $U_k=v_k-\epsilon_k$, $k\in\mathcal{T}$. Since $\arg\max_{k\in\mathcal{T}}\left(v_k-\varepsilon_k\right)=\arg\min_{k\in\mathcal{T}}\left(-v_k+\varepsilon_k\right)$, the best choice probability in the reverse model, denoted by $\widetilde{B}_{\mathcal{T}}\left(i\right)$, $i\in\mathcal{T}$, corresponds to the worst choice probability in the standard model, shifting the deterministic parts from v_k to $-v_k$, $k\in\mathcal{T}$. Similar arguments hold for the worst choice probabilities denoted by $\widetilde{W}_{\mathcal{T}}\left(i\right)$.

Therefore, adapting Eq. (16) of Proposition 1, the best choice probabilities for the reverse RMNL are given by the reverse MNL (RMNL) formula

$$\widetilde{B}_{\mathcal{T}}\left(i\right)=\sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\frac{e^{-v_i}}{\sum_{k\in\mathcal{S}}e^{-v_k}},\ i\in\mathcal{T}. \tag{21}$$

Likewise, the worst choice probabilities have a simple logit form and are given by the worst reverse MNL (WRMNL) formula

$$\widetilde{W}_{\mathcal{T}}\left(i\right)=\frac{e^{-v_i}}{\sum_{k\in\mathcal{T}}e^{-v_k}},i\in\mathcal{T}. \tag{22}$$

The above expression has often been used erroneously in the literature as a surrogate for the worst choice probability (of the MNL). Moreover, using Eq. (19) in Proposition 3, the expected maximum for the RMNL is

$$\mathbb{E}\left[\max_{k\in\mathcal{T}}\left(v_k-\varepsilon_k\right)\right]=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\ln\sum_{k\in\mathcal{S}}e^{-v_k}. \tag{23}$$

On the other hand, the expected minimum benchmark for the RMNL has a logsum form

$$\mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k-\varepsilon_k\right)\right]=\ln\sum_{k\in\mathcal{T}}e^{-v_k}. \tag{24}$$

6.3 Block-Marschak polynomials

Block and Marschak (1960) have proved that a system of best choice probabilities $B_{\mathcal{S}}(i)$, $i\in\mathcal{S}\subset\mathcal{T}$, induced by a probability distribution $P(\cdot)$ on the set of rankings[7] Ω verify conditions on some polynomials $K_{\mathcal{R}}^{B}(i)$, known as the Block-Marschak polynomials. They are defined by

$$K_{\mathcal{R}}^{B}(i)\equiv\sum_{\mathcal{R}\subset\mathcal{S}\subset\mathcal{T}}(-1)^{|\mathcal{S}|-|\mathcal{R}|}B_{\mathcal{S}}(i),\ i\in\mathcal{R}\subset\mathcal{T}, \tag{25}$$

where the superscript B refers to the best choice (implicitly considered in the literature).

Falmagne (1978) has proved that a system of best choice probabilities (with $\sum_{i\in\mathcal{S}}B_{\mathcal{S}}(i)=1$, $\mathcal{S}\subset\mathcal{T}$), verifying the $K_{\mathcal{R}}^{B}(i)\geq 0$, $i\in\mathcal{R}\subset\mathcal{T}$, can be induced by a probability $P(\cdot)$ on Ω.

By analogy, we define the *worst* Block-Marschak polynomials associated to a system of worst choice probabilities $W_{\mathcal{S}}(i)$, $i\in\mathcal{S}\subset\mathcal{T}$, as

$$K_{\mathcal{R}}^{W}(i)\equiv\sum_{\mathcal{R}\subset\mathcal{S}\subset\mathcal{T}}(-1)^{|\mathcal{S}|-|\mathcal{R}|}W_{\mathcal{S}}(i),\ i\in\mathcal{R}\subset\mathcal{T}, \tag{26}$$

where

$$\begin{aligned}
&K_{\mathcal{T}}^{W}(i)=W_{\mathcal{T}}(i);\\
&K_{\mathcal{T}-\{j\}}^{W}(i)=W_{\mathcal{T}-\{j\}}(i)-W_{\mathcal{T}}(i);\\
&K_{\mathcal{T}-\{j,k\}}^{W}(i)=\left[W_{\mathcal{T}-\{j,k\}}(i)-W_{\mathcal{T}-\{j\}}(i)\right]-\left[W_{\mathcal{T}-\{k\}}(i)-W_{\mathcal{T}}(i)\right];\\
&\text{etc.}
\end{aligned}$$

[7] This means that there esists $P(\cdot)$ such that $B_{\mathcal{S}}(i)=\sum_{\rho\in\Omega:\ i\succeq k,\ k\in\mathcal{S}}P(\rho)$, $i\in\mathcal{S}\subset\mathcal{T}$.

The relation between the best and the worst Block-Marschak polynomials are given in:

Lemma 2 *Consider best and worst choice probabilities induced by a probability distribution $P(\cdot)$ on Ω. Their associated best and worst Block-Marschak polynomials obey the following identity:*

$$K_{\mathcal{R}}^{W}(i) = K_{\{i\}\cup(\mathcal{T}-\mathcal{R})}^{B}(i), \; i \in \mathcal{R} \subset \mathcal{T}.$$

Proof. See Appendix B.

Falmagne's theorem can be extended as follows to the worst choice probabilities:

Proposition 4 *Consider a system of worst choice probabilities $W_{\mathcal{S}}(i)$, $i \in \mathcal{S} \subset \mathcal{T}$ (with $\sum_{i\in\mathcal{S}} W_{\mathcal{S}}(i) = 1$, $\mathcal{S} \subset \mathcal{T}$) and assume their worst Block-Marschak polynomials verify $K_{\mathcal{R}}^{W}(i) \geq 0, i \in \mathcal{R} \subset \mathcal{T}$. Than, these probabilities can be induced by a probability $P(\cdot)$ on the set of rankings Ω.*

Proof. See Appendix A.

It is well known that when the probability distribution on the set of rankings is well defined, there exists some random utilities U_i, which satisfy: $B_{\mathcal{S}}(i) = \Pr(U_i \geq U_k, \; \forall k \in \mathcal{S})$, $i \in \mathcal{S} \subset \mathcal{T}$.

6.4 A CES for the worst alternative

We now consider a population of statistically homogeneous and independent consumers. Each consumer is endowed with income y and faces a budget constraint $\sum_{k\in\mathcal{T}} p_k q_k = y$, where p_k is the price and q_k the quantity of good

k, $k \in \mathcal{T}$. By construction, each consumer purchases a single good. Let $U_k = \lambda \ln(y/p_k) + \varepsilon_k$, $k \in \mathcal{T}$, $\lambda > 0$. The agent conditional demand for good i verifies $q_i = y/p_i$, $i \in \mathcal{T}$. Given a subset of goods $\mathcal{S}$, the agent expected (best) demand for good i is: $Q^B_{\mathcal{S}}(i) = B_{\mathcal{S}}(i) \times y/p_i$, $i \in \mathcal{S} \subset \mathcal{T}$. Assuming the errors are i.i.d. Gumbel, we get the MNL formula (15) for $B_{\mathcal{S}}(i)$, and obtain a CES functional form as follows (see Anderson, de Palma and Thisse, 1992, who derived the relationship between the MNL and the CES representative consumer model):

$$Q^B_{\mathcal{S}}(i) = \frac{p_i^{-\lambda}}{\sum_{k \in \mathcal{S}} p_k^{-\lambda}} \frac{y}{p_i}, \ i \in \mathcal{S} \subset \mathcal{T}. \tag{27}$$

Using our analysis of worst choice, we obtain:

Proposition 5 *Consider a finite population of statistically independent consumers, with conditional utility $U_k = \lambda \ln(y/p_k) + \epsilon_k$, $k \in \mathcal{T}$, where ϵ_k are i.i.d. Gumbel. The expected worst demand is given by the worst CES (WCES) formula:*

$$Q^W_{\mathcal{T}}(i) = \left[\sum_{\{i\} \subset \mathcal{S} \subset \mathcal{T}} (-1)^{|\mathcal{S}|-1} \frac{p_i^{-\lambda}}{\sum_{k \in \mathcal{S}} p_k^{-\lambda}} \right] \frac{y}{p_i}, \ i \in \mathcal{T}. \tag{28}$$

Proof. See Appendix A.

Proposition 5, which introduces the WCES shed some light on the social preference of the CES representative consumer. It provides some quantitative guidance concerning the social desirability of the worst choice. For example, when deciding the location of an infrastructure, it matters to know how the consumers benefit if they cannot have access to their best choice and are allocated to their least desirable available one (e.g. a consumer may be allocated to the worst bus line, the least desirable school, the worst hospital or the worst bank branch).This provide useful benchmark in welfare analysis.

7 Estimation

During the last few years, several researchers (Marley and Louviere, 2005; marley, 2012) have shown empirically that elicitation of preferences can be improved by asking respondents to reveal their worst choice or both the best and worst choices instead of their best. We provide here a method to compare the econometric efficiency when asking for the best or the worst alternative using a simple example.

Consider the case of estimating the parameters of an ARUM model where the systematic parts of the utilities are linear in $\boldsymbol{\beta}$, the $(L, 1)$ vector of unknown parameters

$$v_k = \mathbf{z}_k \boldsymbol{\beta}, \; k \in \mathcal{T}, \tag{29}$$

where $\mathbf{z}_k$ denotes the $(1, L)$ vector of characteristics of alternative k. Hereafter, for any $(1, L)$ vector $\mathbf{z}$, we denote by $\mathbf{z}^2$ the symmetric matrix $\mathbf{z}^2 \equiv \mathbf{z}'\mathbf{z}$, where $\mathbf{z}'$ is the transpose of the $\mathbf{z}$ vector.

7.1 Best choice observed for the MNL

Given that the best choice is observed in a (sub)set S$\subset \mathcal{T}$, the contribution to the log-likelihood from a single observation, denoted by $\mathbb{L}_{\mathcal{S}}^{B}$, is given for the MNL model by (see, e.g., Amemiya, 1985)

$$\mathbb{L}_{\mathcal{S}}^{B} = \sum_{i \in \mathcal{S}} \chi_{\mathcal{S}}^{B}(i) \, \mathbf{z}_i \boldsymbol{\beta} - \ln \sum_{i \in \mathcal{S}} e^{\mathbf{z}_i \boldsymbol{\beta}}, \, \mathcal{S} \subset \mathcal{T}, \tag{30}$$

where $\chi_{\mathcal{S}}^{B}(i)$ are the best choice indicators.

The log-concavity of $B_{\mathcal{S}}(i)$ in the vector of utilities v_k, $k \in \mathcal{S}$, which have a linear specification in $\boldsymbol{\beta}$, insures the concavity of $\mathbb{L}_{\mathcal{S}}^{B}$ in $\boldsymbol{\beta}$. Therefore the concavity of the sample log-likelihood is guaranteed if independent observations are considered. A single observation provides a semi-definite positive Fisher information matrix

$$\mathbb{I}_{\mathcal{S}}^{B} \equiv -\mathbb{E}\left(\partial^{2}\mathbb{L}_{\mathcal{S}}^{B}/\partial\boldsymbol{\beta}^{2}\right) = \sum_{i\in\mathcal{S}} B_{\mathcal{S}}\left(i\right)\left(\mathbf{z}_{i} - \mathbf{m}_{\mathcal{S}}^{B}\right)^{2}, \; \mathcal{S}\subset\mathcal{T}, \tag{31}$$

where

$$\mathbf{m}_{\mathcal{S}}^{B} \equiv \sum_{i\in\mathcal{S}} B_{\mathcal{S}}\left(i\right)\mathbf{z}_{i}, \; \mathcal{S}\subset\mathcal{T}. \tag{32}$$

7.2 Worst choice observed for WMNL

Assume that the worst choice in $\mathcal{T}$ is observed. The contribution to the log-likelihood from a single observation, $\mathbb{L}_{\mathcal{T}}^{W}$, is given in the WMNL model by

$$\mathbb{L}_{\mathcal{T}}^{W} = \sum_{i\in\mathcal{T}} \chi_{\mathcal{T}}^{W}\left(i\right)\mathbf{z}_{i}\boldsymbol{\beta} + \sum_{i\in\mathcal{T}} \chi_{\mathcal{T}}^{W}\left(i\right)\ln \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} \frac{\left(-1\right)^{|\mathcal{S}|-1}}{\sum_{k\in\mathcal{S}} e^{\mathbf{z}_{k}\boldsymbol{\beta}}}, \tag{33}$$

where $\chi_{\mathcal{T}}^{W}\left(i\right)$ are the worst choice indicators. We have:

Proposition 6 *For the MNL, the log-likelihood of a single observation of the worst choice in $\mathcal{T}$ is concave and has an information matrix given by:*

$$\mathbb{I}_{\mathcal{T}}^{W} = \sum_{i\in\mathcal{T}} W_{\mathcal{T}}\left(i\right)\left(\mathbf{z}_{i} - \mathbf{m}_{\mathcal{T}}^{W}\left(i\right)\right)^{2}, \tag{34}$$

where

$$\mathbf{m}_{\mathcal{T}}^{W}\left(i\right) \equiv \frac{1}{W_{\mathcal{T}}\left(i\right)} \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} \left(-1\right)^{|\mathcal{S}|-1} B_{\mathcal{S}}\left(i\right)\mathbf{m}_{\mathcal{S}}^{B}, \; i\in\mathcal{T}. \tag{35}$$

Proof. See Appendix A.

7.3 Worst can be better than best

Consider now a particular case involving two alternatives. According to the uniform expansion, each alternative is replicated m times. Consider an individual facing a choice set consisting in a cup of coffee and a cup of tea. According to Yellott (1977), the probability that an individual facing a cup of coffee and a cup of tea is the same as if he were facing m cups of coffee and m cups of tea iff the error terms are i.i.d. Gumbel.[8]

Here we consider a *non-uniform expansion*, where alternative 1 is replicated n_1 times and alternative 2 is replicated n_2 times. There are now two subsets of alternatives denoted by $\mathcal{T}_1$ and $\mathcal{T}_2$ containing each statistical identical alternatives. Varying n_1 and n_2, the probability that the best (or worst) choice belongs to $\mathcal{T}_1$ varies from 0 to 1. We will see in an example that changing n_1 and n_2 allows one to achieve optimal econometric efficiency.

Assume that the utility of an alternative k belonging to $\mathcal{T}_g$ is $U_k = v_g + \varepsilon_k$, $g = 1, 2$. All replicated alternatives have the same systematic part and their random error terms are i.i.d. Without loss of generality, we set $v_1 = \beta \geq 0$ and $v_2 = 0$. The parameter β will be estimated by asking respondents to select their best or worst choice. The probability that the individual selects his best alternative in $\mathcal{T}_1$ is

$$B_{n_1,n_2} \equiv \Pr\left(\beta + \max_{i\in\mathcal{T}_1} \varepsilon_i \geq \max_{i\in\mathcal{T}_2} \varepsilon_i\right). \tag{36}$$

Let $\Phi(.)$ be the CDF and $\phi(.)$ the PDF of ε_i, $i \in \mathcal{T}_1 \cup \mathcal{T}_2$. We can rewrite

[8] Our previous analysis shows that this invariance property does not hold for the worst choice.

(36) as a one-dimensional integral

$$B_{n_1,n_2} = n_1 \int_{-\infty}^{\infty} \Phi^{n_2} (\beta + x) \, \Phi^{n_1 - 1} (x) \, \phi (x) \, dx. \tag{37}$$

The probability $W_{n_1,n_2} \equiv \Pr \left(\beta + \min_{i \in \mathcal{T}_1} \varepsilon_i \leq \min_{i \in \mathcal{T}_2} \varepsilon_i \right)$ that the individual selects his worst alternative in $\mathcal{T}_1$ can be be written as a function of best choice probabilities (see Corollary 1).

We now examine the properties of the estimators $\widehat{\beta}$ of β. If the best choice is observed, the loglikelihood of a single observation is

$$\mathbb{L}^B_{n_1,n_2} = \chi^B_{n_1,n_2} \ln B_{n_1,n_2} + \left(1 - \chi^B_{n_1,n_2} \right) \ln \left(1 - B_{n_1,n_2} \right), \tag{38}$$

where $\chi^B_{n_1,n_2}$ is group $\mathcal{T}_1$ best choice indicator. The second-order derivative of $\mathbb{L}_{n_1,n_2}$ is given by

$$\begin{aligned} \frac{\partial^2 \mathbb{L}_{n_1,n_2}}{\partial \beta^2} &= \left[- \frac{\chi^B_{n_1,n_2}}{B^2_{n_1,n_2}} - \frac{1 - \chi^B_{n_1,n_2}}{1 - B_{n_1,n_2}} \right] \left[\frac{\partial B_{n_1,n_2}}{\partial \beta} \right]^2 \\ &\quad + \left[\frac{\chi^B_{n_1,n_2}}{B_{n_1,n_2}} - \frac{1 - \chi^B_{n_1,n_2}}{1 - B_{n_1,n_2}} \right] \frac{\partial^2 B_{n_1,n_2}}{\partial \beta^2}. \end{aligned}$$

Since the expectation of the second term of the RHS is zero, the variance $V^B_{n_1,n_2}$ of $\widehat{\beta}$ when the best choice is observed is

$$V^B_{n_1,n_2} = \left[\mathbb{E} \left(- \frac{\partial^2 \mathbb{L}_{n_1,n_2}}{\partial \beta^2} \right) \right]^{-1} = \frac{B_{n_1,n_2} \left(1 - B_{n_1,n_2} \right)}{\left(\partial B_{n_1,n_2} / \partial \beta \right)^2}. \tag{39}$$

Clearly, if the worst choice is observed, the variance of the estimator of β is given by an analogous formula

$$V^W_{n_1,n_2} = \frac{W_{n_1,n_2} \left(1 - W_{n_1,n_2} \right)}{\left[\partial W_{n_1,n_2} / \partial \beta \right]^2}. \tag{40}$$

In the logit formulation, where the error terms are i.i.d. Gumbel, we have $B_{n_1,n_2} = n_1 e^{\beta} \left(n_1 e^{\beta} + n_2\right)^{-1}$ and $\partial B_{n_1,n_2}/\partial\beta = B_{n_1,n_2}\left(1 - B_{n_1,n_2}\right)$. Note that these formula are invariant under uniform expansion. Thus, we get: $V^B_{n_1,n_2} = \left[B_{n_1,n_2}\left(1 - B_{n_1,n_2}\right)\right]^{-1}$, which is minimized for equal probabilities.

After some simplifications, the worst choice probability is given explicitly by

$$W_{n_1,n_2} \equiv \sum_{q_1=1}^{n_1} \sum_{q_2=0}^{n_2} (-1)^{q_1+q_2-1} \binom{n_1}{q_1} \binom{n_2}{q_2} \frac{q_1 e^{\beta}}{q_1 e^{\beta} + q_2}. \tag{41}$$

This formula easily allows to compute numerically the the variance $V^W_{n_1,n_2}$.

Fig. 1 below shows, for the logit formulation and for $n = 10$, how $V^B_{n_1,n_2}$ behave with respect to β and the distribution of the alternatives among the two groups

INSERT FIGURE 1 ABOUT HERE

It is clear from equations (37) that at $\beta = 0$, the variances coincide when the number of alternatives in each group are swapped. The minimal variance, 4, occurs when the best choice probabilities within the two groups are balanced (see the analytical expressions). When $\beta = 0$, this optimal design happens for $n_1 = 5$. As intuition suggests, the variance at $\beta = 0$ is minimal when the numbers of alternatives in each group are equalized. For $\beta > 0$, the optimal design entails $n_1 \leq n_2$. Fig. 2 is the counterpart of Fig. 1 if the worst choice is observed.

INSERT FIGURE 2 ABOUT HERE

Note that the lowest variance in this case is 0.87 instead of 4 if the best choice is observed. The lower envelope is rapidly increasing with β. For $\beta \in (0, 2.16)$, the optimal design for the WMNL remains lower than the minimum variance for the MNL. Fig. 3 displays the ratio $V^B_{n_1,n_2}/V^W_{n_1,n_2}$ with respect to β for $n = 3, ..., 10$, when the design (distribution of the alternatives among the two groups) is optimized for each choice situation (best or worst).

INSERT FIGURE 3 ABOUT HERE

For example, for 3 alternatives, the ratio for $\beta = 0$ is 1.56 which means that 100 observations if the worst choice is observed provides as much information as 156 observations with the best choice observed. For 10 alternatives and $\beta = 0$, 459 observations are required for the best choice to achieve the same accuracy as with 100 observations with the worst choice.

One may wonder why the efficiencies for the best and worst choices are different. One explanation could be the fact that the Gumbel is an asymmetric distribution with positive skewness. However, the discussion of the probit model shows that this intuition is false.

The probability that the best alternative belongs to $\mathcal{T}_1$ in the reverse model is such that: $\widetilde{B}_{n_1,n_2} \equiv \Pr\left(\beta + \max_{i\in\mathcal{T}_1}(-\varepsilon_i) \geq \max_{i\in\mathcal{T}_2}(-\varepsilon_i)\right)$. Clearly enough, we have $W_{n_1,n_2} = 1 - \widetilde{B}_{n_2,n_1}$. As a consequence, the comparison between the estimates obtained in the standard and the reverse logit model can be deduced from Figure 2.[9]

The same comparison between best and worst can be performed for the probit model. In this case, there is an extra simplification with respect to the logit since the error terms are symmetric. In a symmetric model, since $\widetilde{B}_{n_1,n_2} = B_{n_1,n_2}$, we have $W_{n_1,n_2} = 1 - B_{n_2,n_1}$. Obviously, we have $V^W_{n_1,n_2} = V^B_{n_2,n_1}$. The variance of $\widehat{\beta}$ when the worst choice is observed is equal to the variance of $\widehat{\beta}$ in the reverse model when the numbers of alternatives in each group are swapped. This symmetry property will be used in the numerical results derived for our model.

Note that the symmetry of the error terms does not imply that the variance is the same when elicitating preferences with best or worst choices. This

[9] Misra (2005) has shown that the RMNL performs better than the MNL for a specific application in marketing. He used a GEV model in his analysis.

is because the choice process requires a comparison of the best (or worst) alternative in group 1 and in group 2. As we know, the maximum (or the minimum) of i.i.d. normal distributions is a positively (or negatively for the minimum) skewed distribution, which implies the problem is no longer symmetric. Numerical simulations show that when $\beta > 0$, the variance of the estimator is smaller for the design leading to a choice probability closer to 1/2.

In the general case, for symmetric models, the standard and the reverse best choice probabilities coincide: $\widetilde{B}_{\mathcal{T}}(i) = B_{\mathcal{T}}(i)$. The worst choice probabilities $W_{\mathcal{T}}(i)$ can be obtained by simply using the best choice probabilities $B_{\mathcal{T}}(i)$ by changing v_k into $-v_k$, $k \in \mathcal{T}$. Therefore, worst choice data with a probit model can be estimated using the standard packages programmed, substituting the best choice indicators by the worst choice ones. However, one has to reverse the estimated parameters $\widehat{\boldsymbol{\beta}}$ and taking $-\widehat{\boldsymbol{\beta}}$.

Assume for example that survey data where respondents are asked to reveal their best choice have been estimated with a probit model. If the same survey has to be administered two years later, one may wonder if it would be more efficient to ask respondents to reveal their worst choice. One can proceed as follows: (a) using the initial estimates, generate worst choice alternatives for the population of individual; (b) estimate the parameter and the variance for this new set of worst choice alternatives; (c) compare the loglikelihood of the two models using standard methods. This procedure discussed above in (b) can be performed with standard packages. This provides some grounded reason to ask for best or worst choice in the new survey.

Acknowledgments. The authors are grateful to the participants of the International Choice Modelling Conference 2011 and in seminars in Ecole Polytechnique; National Taiwan University; ETH; University of Virginia; KULeven; Leontiev Center; University of Laval; UBC. We thank S. Anderson, P. Chiappori, C. Holt, J. Louviere, and in particular A. Marley who provide us with warm encouragements.

References

[1] **Anas, Alex and Chu, Chaushie.** "Discrete Choice Models and the Housing Price and Travel to Work Elasticities of Location Demand." *Journal of Urban Economics*, 1984, 15(1), pp. 107–23.

[2] **Anderson, Simon P. and de Palma, André.** "Reverse Discrete Choice Models." *Regional Science and Urban Economics*, 1999, 29(6), 745–64.

[3] **Anderson, Simon P.; de Palma, André and Thisse, Jacques-François.** *Discrete choice theory of product differentiation.* Cambridge, MA: The MIT Press, 1992.

[4] **Ben-Akiva, Moshe and Lerman, Steven R.** *Discrete choice analysis.* Cambridge, MA: The MIT Press, 1985.

[5] **Bender, Edward A. and Goldman, Jay R.** "On the Applications of Möbius Inversion in Combinatorial Analysis." *American Mathematical Monthly*, 1975, 82(8), pp. 789-803, .

[6] **Block, Henry D. and Marschak, Jacob.** "Random Ordering and Stochastic Theories of Responses," in Ingram Olkin et al., eds., *Contributions to probability and statistics.* Stanford, CA: Stanford University Press, 1960, pp. 97-132.

[7] **Caplin, Andrew and Nalebuff, Barry.** "Aggregation and Imperfect Competition: On the Existence of Equilibrium." *Econometrica*, 1991, 59(1), pp. 25-59.

[8] **Chintagunta, Pradeep K. and Nair, Harikesh S.** "Structural Workshop Paper - Discrete-Choice Models of Consumer Demand in Marketing." *Marketing Science*, 2011, 30(6), pp. 977-996.

[9] **de Palma, André and Kilani, Karim.** "Transition Choice Probabilities and Welfare Analysis in Additive Random Utility Models." *Economic Theory*, 2011, 46(3), pp. 427–454.

[10] **Falmagne, Jean-Claude.** "A Representation Theorem for Finite Random Scale Systems." *Journal of Mathematical Psychology*, 1978, 18(1), pp. 52–72.

[11] **Finney, David John.** *Probit Analysis*. Cambridge, UK: Cambridge University Press, 1947.

[12] **Fiorini, Samuel.** "A Short Proof of a Theorem of Falmagne." *Journal of Mathematical Psychology*, 2004, 48(1), pp. 80–82.

[13] **Hoffman, Saul D. and Duncan, Greg J.** "Multinomial and Conditional Logit Discrete-Choice Models in Demography." *Demography*, 1988, 25(3), pp. 415-27. .

[14] **Kokovin, Sergey; Parenti, Mathieu; Thisse, Jacques-François and Zhelobodko, Evgeny.** "Monopolistic Competition: Beyond the CES." *Econometrica*, 2012, 80(6), pp. 2765-84.

[15] **Labeaga, Jose M.; Xisco, Oliver and Spadaro, Amedeo,** . "Discrete Choice Models of Labour Supply, Behavioural Microsimulation and the Spanish Tax Reforms." *Journal of Economic Inequality*, 2008, 6(3), pp. 247-73.

[16] **Luce, Duncan R.** *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley, 1959.

[17] **Manski, Charles F.** "The Structure of Random Utility Models." *Theory and Decision*, 1977, 8(3), pp. 229-54.

[18] **Marley, Antony A. J. and Louviere, Jordan J**. "Some Probabilistic Models of Best, Worst, and Best–Worst choices." *Journal of Mathematical Psychology*, 2005, 49, pp. 464–80.

[19] **Marley, Antony A. J.** "The Best-Worst Method for the Study of Preferences: Theory and Application,' in Peter A. Frensch and Ralf Schwarzer, eds., *Cognition and neuropsychology: International perspectives on psychological science (Vol. 1).* New York, NY: Psychology Press, 2010, pp. 147-58.

[20] **McFadden, Daniel.** . "Conditional logit analysis of qualitative choice behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*, New York, NY: Academic Press, 1974, pp. 105–42.

[21] **McFadden, Daniel.** "Economic Choices." *American Economic Review*, 2001, 91(3), pp. 351-78.

[22] **Misra, Sanjog.** "Generalized Reverse Discrete Choice Models." *Quantitative Marketing and Economics*, 2005, 3(2), pp. 175–200.

[23] **Poincaré, Henri.** *Calcul des probabilités*. Paris: Gauthier-Villars, 1912.

[24] **Rosen, Harvey S. and Small, Kenneth A.** "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, 1981, 49(1), pp. 105-30.

[25] **Thurstone, Louis Leon**. "A Law of Comparative Judgement." *Psychological Review*, 1927, 34, pp. 278-86.

[26] **Train, Kenneth.** *Discrete choice methods with simulation*. Cambridge, MA: Cambridge University Press, 2003.

[27] **Warner, Stanley L.** *Stochastic choice of mode in urban travel: A study in binary choice*. Evanston, Il: Northwestern University Press, 1962.

[28] **Warner, Stanley L.** "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association*, 1965, 60(309), pp. 63-69.

[29] **Yellot, John I.** "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution." *Journal of Mathematical Psychology*, 1977, 15(2), pp. 109-44.

Appendix A: proofs of Propositions

Proof of Proposition 1. The WMNL formula (16) is obtained using Eq. (5) and the MNL formula given by (15). The proof of log-concavity of $W_{\mathcal{T}}(i)$ in relies on its integral form which is given by

$$W_{\mathcal{T}}(i) = \int_{-\infty}^{\infty} e^{v_i - x} \exp\left(e^{v_i - x}\right) \prod_{k \in \mathcal{T} - \{i\}} \left[1 - \exp\left(e^{v_k - x}\right)\right] dx.$$

The integrand is is log-concave in x and in the vector of v_k, $k \in \mathcal{T}$, as a product of log-concave functions. According to the Prékopa–Leindler inequality (see Caplin and Nalebuff, 1991), the integral over x is still log-concave in the vector of v_k, $k \in \mathcal{T}$.

Proof of Proposition 2. Using Eq. (7) and Lemma 1, we obtain that $BW_{\mathcal{T}}(i,j) = B_{\mathcal{T}}(i)\, W_{\mathcal{T}-\{i}(j)$. Then using the MNL formula (15) and the WMNL formula (16), Eq. (17) is obtained.

Proof of Proposition 3 Use (12) and (18) to get

$$\mathbb{E}\left[\min_{k \in \mathcal{T}} (v_k + \varepsilon_k)\right] = \sum_{\mathcal{S} \subset \mathcal{T}} (-1)^{|\mathcal{S}|-1} \ln \sum_{k \in \mathcal{T}} e^{v_k}.$$

Proof of Proposition 4. We associate to system $W_{\mathcal{S}}(i)$, the following system $B_{\mathcal{S}}(i) \equiv \sum_{\{i\} \subset \mathcal{U} \subset \mathcal{S}} (-1)^{|\mathcal{U}|-1} W_{\mathcal{U}}(i)$, $i \in \mathcal{S} \subset \mathcal{T}$. We need to prove first that this is indeed a system of best choice probabilities. We have

$$\sum_{i \in \mathcal{S}} B_{\mathcal{S}}(i) = \sum_{i \in \mathcal{S}} \sum_{\{i\} \subset \mathcal{U} \subset \mathcal{S}} (-1)^{|\mathcal{U}|-1} W_{\mathcal{U}}(i).$$

Inverting the two sum signs of the RHS, we get

$$\sum_{i\in\mathcal{S}} B_{\mathcal{S}}(i) = \sum_{\mathcal{U}\subset\mathcal{S}} (-1)^{|\mathcal{U}|-1} \sum_{i\in\mathcal{U}} W_{\mathcal{U}}(i).$$

Since $\sum_{i\in\mathcal{U}} W_{\mathcal{U}}(i) = 1$, we obtain $\sum_{i\in\mathcal{S}} B_{\mathcal{S}}(i) = \sum_{\mathcal{U}\subset\mathcal{S}} (-1)^{|\mathcal{U}|-1} = 1$. We proof that $B_{\mathcal{S}}(i) \geq 0$, showing that $B_{\{i\}\cup(\mathcal{T}/\mathcal{S})}(i) = \sum_{\{i\}\subset\mathcal{R}\subset\mathcal{R}} K_{\mathcal{R}}^{W}(i)$. We have

$$\sum_{\{i\}\subset\mathcal{R}\subset\mathcal{S}} K_{\mathcal{R}}^{W}(i) = \sum_{\{i\}\subset\mathcal{R}\subset\mathcal{S}} \sum_{\mathcal{R}\subset\mathcal{U}\subset\mathcal{T}} (-1)^{|\mathcal{U}|-|\mathcal{R}|} W_{\mathcal{U}}^{W}(i),$$

which can be rewritten as

$$\sum_{\{i\}\subset\mathcal{R}\subset\mathcal{S}} K_{\mathcal{R}}^{W}(i) = \sum_{\{i\}\subset\mathcal{U}\subset\mathcal{T}} (-1)^{|\mathcal{U}|} \left[\sum_{\mathcal{R}\subset\mathcal{U}\cap\mathcal{S}} (-1)^{|\mathcal{R}|}\right] W_{\mathcal{U}}^{W}(i).$$

From the binomial formula, the term into bracket is -1 if $\mathcal{U}\cap\mathcal{S} = \{i\}$ and zero otherwise, so that

$$\sum_{\{i\}\subset\mathcal{R}\subset\mathcal{S}} K_{\mathcal{R}}^{W}(i) = \sum_{\{i\}\subset\mathcal{U}\subset\{i\}\cup(\mathcal{T}/\mathcal{S})} (-1)^{|\mathcal{U}|-1} W_{\mathcal{U}}^{W}(i) = B_{\{i\}\cup(\mathcal{T}/\mathcal{S})}(i).$$

Therefore, the fact that $K_{\mathcal{R}}^{W}(i) \geq 0$, $i \in \mathcal{R} \subset \mathcal{T}$, insures that $B_{\mathcal{S}}(i) \geq 0$, $i \in \mathcal{S} \subset \mathcal{T}$, so that the best choice probability system is well defined. Now, according to Lemma , the best Block-Marshack polynomials $K_{\mathcal{R}}^{B}(i)$ are positive. According to Falmagne (1978, Theorem 4), there exists a probability $P(\cdot)$ defined on Ω.

Proof of Proposition 5. Since $Q_{\mathcal{T}}^{W}(i) = W_{\mathcal{T}}(i)(y/p_i)$, $i \in \mathcal{T}$, using Theorem 1, we get

$$Q_{\mathcal{T}}^{W}(i) = \left[\sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} B_{\mathcal{S}}(i)\right] \frac{y}{p_i},\ i \in \mathcal{T},$$

which can be rewritten as $Q_{\mathcal{T}}^{W}(i) = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} Q_{\mathcal{S}}^{B}(i)$, $i \in \mathcal{T}$. Finally, using Eq. (27), we get the required form given by Eq. (28).

Proof of Proposition 6. Since $\ln B_{\mathcal{S}}(i) = \mathbf{z}_i\boldsymbol{\beta} - \ln \sum_{k\in\mathcal{S}} e^{\mathbf{z}_k\boldsymbol{\beta}}$, we get

$$\frac{\partial \ln B_{\mathcal{S}}(i)}{\partial \boldsymbol{\beta}} = \mathbf{z}_i - \sum_{k\in\mathcal{S}} B_{\mathcal{S}}(k)\,\mathbf{z}_k = \mathbf{z}_i - \mathbf{m}_{\mathcal{S}}^{B},$$

where $\mathbf{m}_{\mathcal{S}}^{B}$ is defined by (32). Therefore, the derivative of the worst choice probability is

$$\frac{\partial W_{\mathcal{T}}(i)}{\partial \boldsymbol{\beta}} = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} B_{\mathcal{S}}(i)\left(\mathbf{z}_i - \mathbf{m}_{\mathcal{S}}^{B}\right).$$

Expanding the RHS sum of above equation, we get

$$\frac{\partial W_{\mathcal{T}}(i)}{\partial \boldsymbol{\beta}} = W_{\mathcal{T}}(i)\,\mathbf{z}_i - \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} B_{\mathcal{S}}(i)\,\mathbf{m}_{\mathcal{S}}^{B}.$$

Using the definition of $\mathbf{m}_{\mathcal{T}}^{W}(i)$ given by (35), we obtain

$$\frac{\partial \ln W_{\mathcal{T}}(i)}{\partial \boldsymbol{\beta}} = \mathbf{z}_i - \mathbf{m}_{\mathcal{T}}^{W}(i), \tag{42}$$

so that

$$\frac{\partial \mathbb{L}_{\mathcal{T}}^{W}}{\partial \boldsymbol{\beta}} = \sum_{i\in\mathcal{T}} \overline{\chi}_{\mathcal{T}}(i)\,\mathbf{z}_i - \sum_{i\in\mathcal{T}} \overline{\chi}_{\mathcal{T}}(i)\,\mathbf{m}_{\mathcal{T}}^{W}(i).$$

Differentiation of above equation yields the Hessian matrix

$$\frac{\partial^2 \mathbb{L}_{\mathcal{T}}^{W}}{\partial \boldsymbol{\beta}^2} = -\sum_{i\in\mathcal{T}} \overline{\chi}_{\mathcal{T}}(i)\,\frac{\partial \mathbf{m}_{\mathcal{T}}^{W}(i)}{\partial \boldsymbol{\beta}}.$$

Applying the expectation operator and using the fact that $\mathbb{E}\left[\overline{\chi}_{\mathcal{T}}(i)\right] = W_{\mathcal{T}}(i)$, we obtain the information matrix

$$\mathbb{I}_{\mathcal{T}}^{W}=\sum_{i\in\mathcal{T}}W_{\mathcal{T}}\left(i\right)\frac{\partial\mathbf{m}_{\mathcal{T}}^{W}\left(i\right)}{\partial\boldsymbol{\beta}}.$$

Now let $\mathbf{m}_{\mathcal{T}}^{W}\equiv\sum_{i\in\mathcal{T}}W_{\mathcal{T}}\left(i\right)\mathbf{m}_{\mathcal{T}}^{W}\left(i\right)$. We can rewrite above equation as

$$\mathbb{I}_{\mathcal{T}}^{W}=\frac{\partial\mathbf{m}_{\mathcal{T}}^{W}}{\partial\boldsymbol{\beta}}-\sum_{i\in\mathcal{T}}\frac{\partial W_{\mathcal{T}}\left(i\right)}{\partial\boldsymbol{\beta}}\left(\mathbf{m}_{\mathcal{T}}^{W}\left(i\right)\right)'. \tag{43}$$

It is worth to notice that $\mathbf{m}_{\mathcal{T}}^{W}$ has an alternative expression $\mathbf{m}_{\mathcal{T}}^{W}\equiv\sum_{i\in\mathcal{T}}W_{\mathcal{T}}\left(i\right)z_{i}$. Indeed, we have

$$\mathbf{m}_{\mathcal{T}}^{W}=\sum_{i\in\mathcal{T}}\sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}B_{\mathcal{S}}\left(i\right)\mathbf{m}_{\mathcal{S}}^{B}.$$

Permuting the two sum signs, we get

$$\mathbf{m}_{\mathcal{T}}^{W}=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\sum_{i\in\mathcal{S}}B_{\mathcal{S}}\left(i\right)\mathbf{m}_{\mathcal{S}}^{B}.$$

Since $\sum_{i\in\mathcal{S}}B_{\mathcal{S}}\left(i\right)=1$, we obtain the following form $\mathbf{m}_{\mathcal{T}}^{W}=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\mathbf{m}_{\mathcal{S}}^{B}$. Now, using the definition of $\mathbf{m}_{\mathcal{S}}^{B}$, we have

$$\mathbf{m}_{\mathcal{T}}^{W}=\sum_{i\in\mathcal{T}}\sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}B_{\mathcal{S}}\left(i\right)\mathbf{z}_{i},$$

yielding the required form $\mathbf{m}_{\mathcal{T}}^{W}=\sum_{i\in\mathcal{T}}W_{\mathcal{T}}\left(i\right)\mathbf{z}_{i}$. Therefore, using (43) and differentiating the above equation, we obtain

$$\mathbb{I}_{\mathcal{T}}^{W}=\sum_{i\in\mathcal{T}}\frac{\partial W_{\mathcal{T}}\left(i\right)}{\partial\boldsymbol{\beta}}\mathbf{z}_{i}'-\sum_{i\in\mathcal{T}}\frac{\partial W_{\mathcal{T}}\left(i\right)}{\partial\boldsymbol{\beta}}\left(\mathbf{m}_{\mathcal{T}}^{W}\left(i\right)\right)',$$

which can be rewritten as

$$\mathbb{I}_{\mathcal{T}}^{W} = \sum_{i\in\mathcal{T}} \frac{\partial W_{\mathcal{T}}(i)}{\partial \boldsymbol{\beta}} \left(\mathbf{z}_i - \mathbf{m}_{\mathcal{T}}^{W}(i)\right)'.$$

Therefore, using (42), we get

$$\mathbb{I}_{\mathcal{T}}^{W} = \sum_{i\in\mathcal{T}} W_{\mathcal{T}}(i) \left(\mathbf{z}_i - \mathbf{m}_{\mathcal{T}}^{W}(i)\right)^2.$$

which is Eq. (34).

Appendix B: Proof of Corollaries and Lemma

Proof of Corollary 1 Dividing both sides of equation (1), we get

$$\frac{W_{\mathcal{T}}^{\#}(i)}{N} = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \frac{B_{\mathcal{S}}^{\#}(i)}{N}.$$

Therefore

$$\lim_{N\to\infty} \frac{W_{\mathcal{T}}^{\#}(i)}{N} = \lim_{N\to\infty} \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \frac{B_{\mathcal{S}}^{\#}(i)}{N}.$$

Inverting the limit and the sum signs, we obtain

$$\lim_{N\to\infty} \frac{W_{\mathcal{T}}^{\#}(i)}{N} = \sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \lim_{N\to\infty} \frac{B_{\mathcal{S}}^{\#}(i)}{N}.$$

Since $\lim_{N\to\infty} W_{\mathcal{T}}^{\#}(i)/N = \overline{\mathbb{P}}_i^{\mathcal{T}}$, using (4), the required identity follows.

Proof of Corollary 2 We first derive both sides of Identity (12)

$$\frac{\partial \mathbb{E}\left[\min_{k\in\mathcal{T}}(v_k+\varepsilon_k)\right]}{\partial v_i} = \frac{\partial}{\partial v_i} \sum_{\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-1} \mathbb{E}\left[\max_{k\in\mathcal{S}}(v_k+\varepsilon_k)\right];$$

which can be rewritten as

$$\frac{\partial \mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]}{\partial v_i}=\sum_{\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}\frac{\partial \mathbb{E}\left[\max_{k\in\mathcal{S}}\left(v_k+\varepsilon_k\right)\right]}{\partial v_i}.$$

Applying the standard Roy's Identity (13) leads

$$\frac{\partial \mathbb{E}\left[\min_{k\in\mathcal{T}}\left(v_k+\varepsilon_k\right)\right]}{\partial v_i}=\sum_{\{i\}\subset\mathcal{S}\subset\mathcal{T}}\left(-1\right)^{|\mathcal{S}|-1}B_{\mathcal{S}}\left(i\right).$$

The right hand side of the above equation is, using (5), equal to $W_{\mathcal{T}}\left(i\right)$.

Proof of Lemma 1. The expectation $\mathbb{E}\left[\chi_{\mathcal{T}}\left(i\right)\chi_{\mathcal{S}}^{B}\left(i\right)\right]$ corresponds to the probability

$$\mathbb{E}\left[\chi_{\mathcal{T}}\left(i\right)\chi_{\mathcal{S}}^{B}\left(j\right)\right]=\Pr\left(U_k\leq U_i,\ k\in\mathcal{T}-\mathcal{S};U_k\leq U_j,\ k\in\mathcal{S}\right).$$

In the MNL case, it is given by a double integral (Euler's constant γ is ignored)

$$\mathbb{E}\left[\chi_{\mathcal{T}}\left(i\right)\chi_{\mathcal{S}}^{B}\left(j\right)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{x_i}\frac{e^{v_i-x_i}e^{v_j-x_j}}{\exp\left(\sum_{k\in\mathcal{T}-\mathcal{S}}e^{v_k-x_i}+\sum_{k\in\mathcal{S}}e^{v_k-x_j}\right)}dx_j dx_i.$$

A first integration in x_j yields

$$\mathbb{E}\left[\chi_{\mathcal{T}}\left(i\right)\chi_{\mathcal{S}}^{B}\left(j\right)\right]=\int_{-\infty}^{\infty}e^{v_i-x_i}\frac{e^{v_j}}{\sum_{k\in\mathcal{S}}e^{v_k}}\exp\left(-\sum_{k\in\mathcal{T}}e^{v_k-x_i}\right)dx_i.$$

A second integration gives

$$\mathbb{E}\left[\chi_{\mathcal{T}}\left(i\right)\chi_{\mathcal{S}}^{B}\left(j\right)\right]=\frac{e^{v_i}}{\sum_{k\in\mathcal{T}}e^{v_k}}\frac{e^{v_j}}{\sum_{k\in\mathcal{S}}e^{v_k}},$$

which coincides with $B_{\mathcal{T}}\left(i\right)B_{\mathcal{S}}\left(j\right)$, yielding $c_{\mathcal{S}}\left(i,j\right)=0$.

Proof of Lemma 2. Using (1) and (26), we obtain

$$K_{\mathcal{R}}^{W}(i) = \sum_{\mathcal{R}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|-|\mathcal{R}|} \sum_{\{i\}\subset\mathcal{Q}\subset\mathcal{S}} (-1)^{|\mathcal{Q}|-1} B_{\mathcal{Q}}(i) .$$

Inversion of the two sum signs yields

$$K_{\mathcal{R}}(i) = \sum_{\{i\}\subset\mathcal{Q}\subset\mathcal{T}} (-1)^{|\mathcal{Q}|-1-|\mathcal{R}|} \left[\sum_{\mathcal{R}\cup\mathcal{Q}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|} \right] B_{\mathcal{Q}}(i) .$$

Notice that

$$\begin{aligned} \sum_{\mathcal{R}\cup\mathcal{Q}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|} &= \sum_{\emptyset\subseteq\mathcal{S}'\subset\mathcal{T}-(\mathcal{R}\cup\mathcal{Q})} (-1)^{|\mathcal{S}'|} \\ &= \sum_{q=0}^{|\mathcal{T}|-|\mathcal{R}\cup\mathcal{Q}|} \binom{|\mathcal{T}| - |\mathcal{R}\cup\mathcal{Q}|}{q} (-1)^{|\mathcal{S}|} . \end{aligned}$$

Using the binomial expansion formula, we obtain

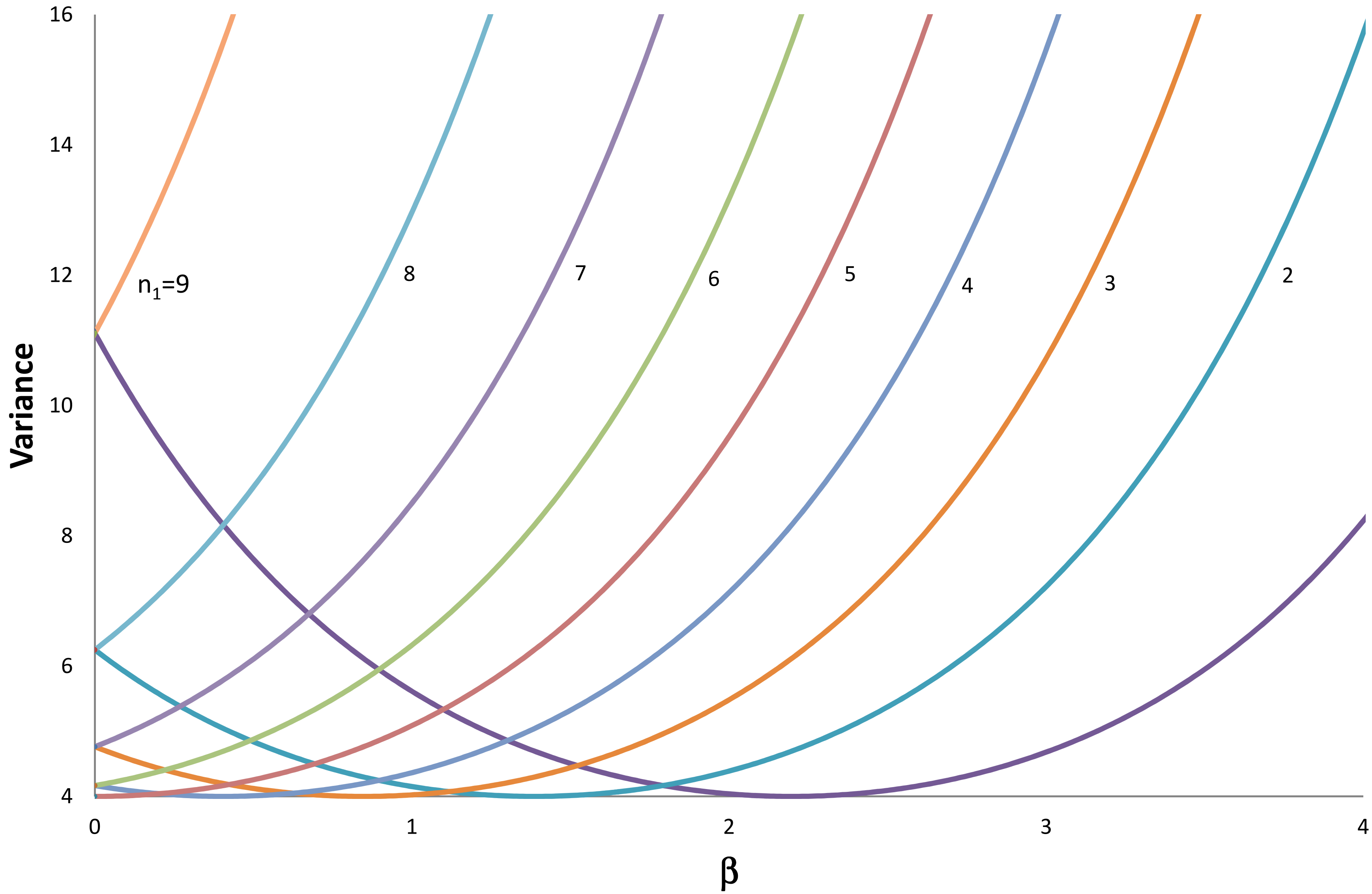
$$\sum_{\mathcal{R}\cup\mathcal{Q}\subset\mathcal{S}\subset\mathcal{T}} (-1)^{|\mathcal{S}|} = \begin{cases} (-1)^{|\mathcal{T}|} & \text{if } \mathcal{R}\cup\mathcal{Q}=\mathcal{T} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$K_{\mathcal{R}}^{W}(i) = \sum_{\{i\}\subset\mathcal{Q}\subset\{i\}\cup(\mathcal{T}\backslash\mathcal{R})} (-1)^{|\mathcal{Q}|-(|\mathcal{T}|-|\mathcal{R}|+1)} W_{\mathcal{Q}}(i) ,$$

yielding $K_{\mathcal{R}}^{W}(i) = K_{\{i\}\cup(\mathcal{T}\backslash\mathcal{R})}^{B}(i)$.

Fig. 1: Variance with respect to β and n_1 for the MNL (n=10)



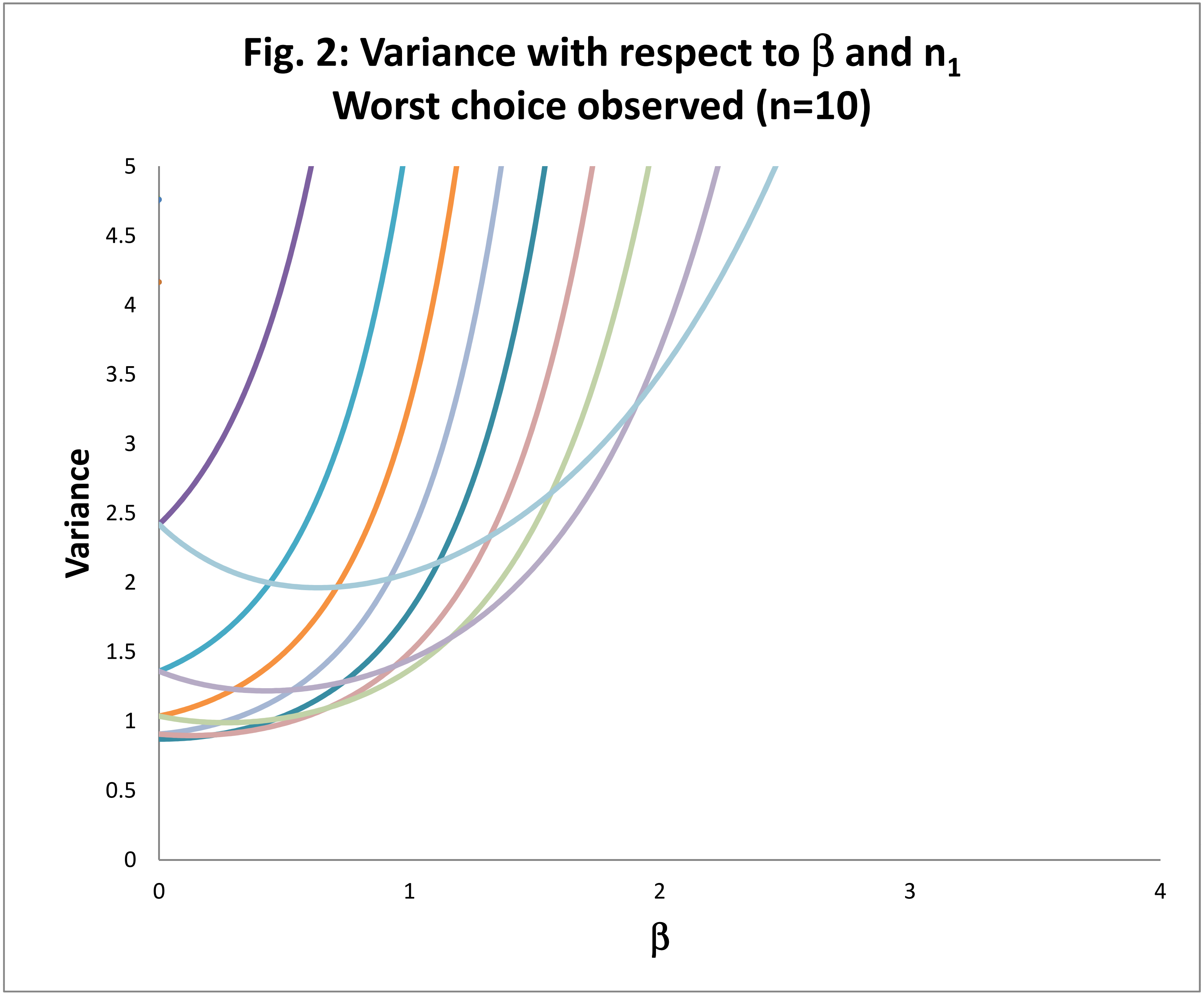


Fig. 2: Variance with respect to β and n1
Worst choice observed (n=10)
Variance
β
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
1
2
3
4

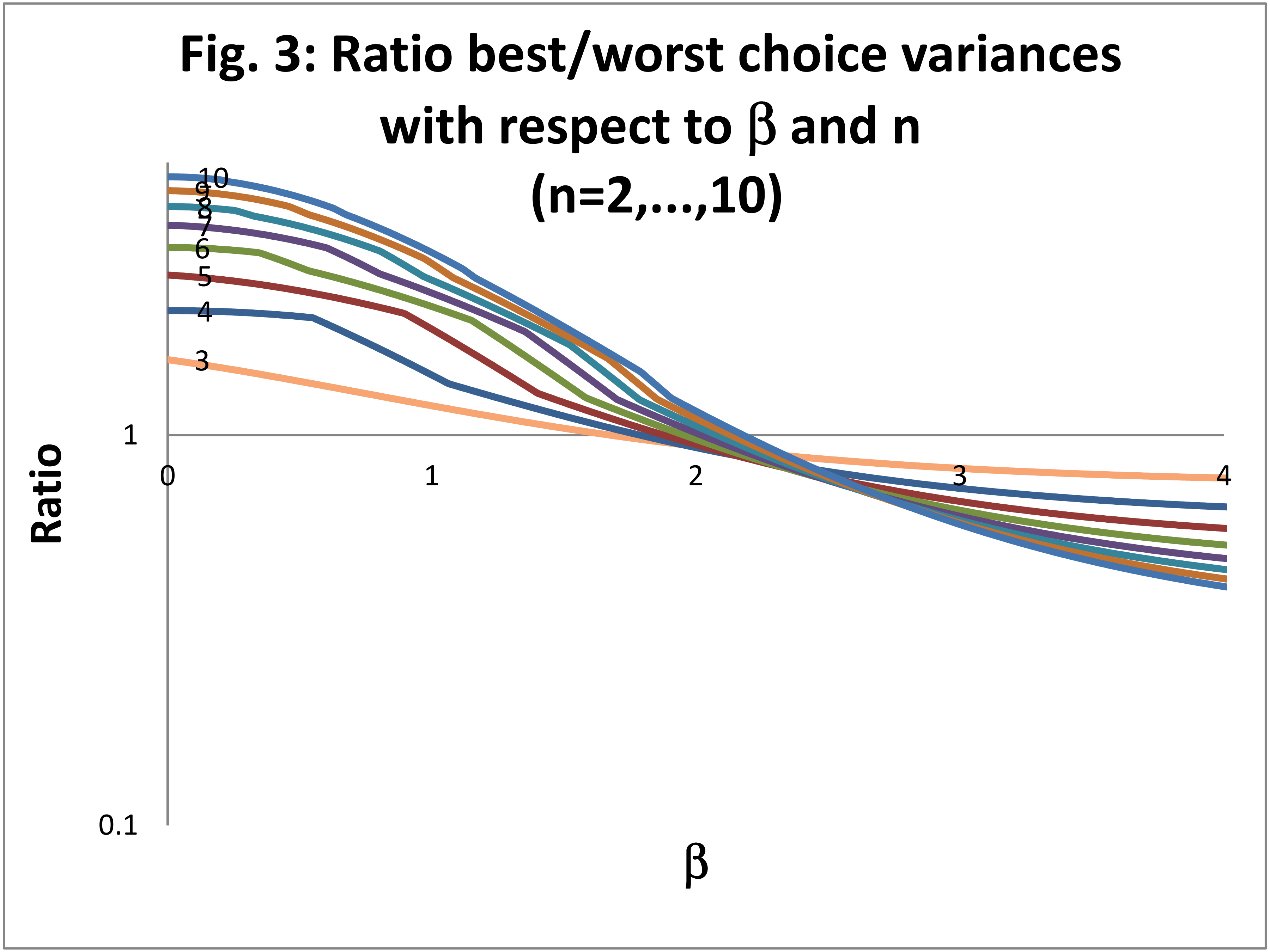


Fig. 3: Ratio best/worst choice variances
with respect to β and n
(n=2,...,10)
10
9
8
7
6
5
4
3
1
0.1
0
1
2
3
4
Ratio
β