Estimation of Logit and Probit models using best, worst and best-worst choices

Paolo Delle Site, Karim Kilani, Valerio Gatta, Edoardo Marcucci, André de Palma

To cite this version:

Paolo Delle Site, Karim Kilani, Valerio Gatta, Edoardo Marcucci, André de Palma. Estimation of Logit and Probit models using best, worst and best-worst choices. 2018. hal-01953581

HAL Id: hal-01953581
https://hal.science/hal-01953581

Preprint submitted on 13 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of Logit and Probit models using best, worst and best-worst choices

Paolo Delle Site[a,*], Karim Kilani[b], Valerio Gatta[c], Edoardo Marcucci[d], André de Palma[e]

[a] *University Niccolò Cusano, Rome, Italy*
[b] *CNAM LIRSA, Paris, France*
[c] *University Roma Tre, Rome, Italy*
[d] *Molde University College, Molde, Norway*
[e] *ENS Paris Saclay, France*

Abstract

The paper considers models for best, worst and best-worst choice probabilities, that use a single common set of random utilities. Choice probabilities are derived for two distributions of the random terms: i.i.d. extreme value, i.e. Logit, and multivariate normal, i.e. Probit. In Logit, best, worst and best-worst choice probabilities have a closed form. In Probit, worst choice probabilities are simply obtained from best choice probabilities by changing the sign of the systematic utilities. Strict log-concavity of the likelihood, with respect to the coefficients of the systematic utilities, holds, under a mild necessary and sufficient condition of absence of perfect multicollinearity in the matrix of alternative and individual characteristics, for best, worst and best-worst choice probabilities in Logit, and for best and worst choice probabilities in Probit. The assumption of substitutability between best and worst choices is tested with data on mode choice, collected for the assessment of user responses to urban congestion charging policies. The numerical results suggest significantly different preferences between best and worst choices, even accounting for scale differences, in both Logit and Probit models. Worst choice data exhibit coefficient attenuation, less pronounced in Probit than in Logit, and higher mean values of travel time savings with larger confidence intervals.

Keywords: Best-worst choices, Congestion charge, Logit, Probit, Random utility model, Strict log-concavity

*Corresponding author

Email addresses: paolo.dellesite@unicusano.it (Paolo Delle Site), karkil2205@gmail.com (Karim Kilani), valerio.gatta@uniroma3.it (Valerio Gatta), edoardo.marcucci@himolde.no (Edoardo Marcucci), andre.de-palma@ens-paris-saclay.fr (André de Palma)

1. Introduction

1.1. Motivation

Researchers use increasingly stated preference data providing partial or full rankings, because they can provide augmented information with respect to methods based on asking the favorite alternative only.

When data on full rankings are available, these are frequently analyzed, for the higher ease of estimation, with the use of the rank-ordered Logit model, introduced by Beggs et al. (1981). An alternative model, of more recent use, is the rank-ordered Probit, introduced by Hajivassiliou & Ruud (1994). Examples in transportation of the rank-ordered Logit are found in Beggs *et al.* (1981) and Ben-Akiva *et al.* (1992). An example of the use of the rank-ordered Probit is in Nair *et al.* (2018a).

However, ranking data can be based on best and worst choices only. Louviere & Islam (2008) suggest that judging items at extremes provides advantages in cognitive efforts on discriminating among items of intermediate importance. In addition, the questionnaire task appears easier for the respondent. Thus, it is also suggested the approach to be of interest to applications concerned with complete rankings, obtained by iterating best-and-worst tasks. Examples in transportation of the use of best and worst data are rare (Beck *et al.*, 2017). The present paper is concerned with this type of partial ranking.

Louviere & Woodworth (1990) and Finn & Louviere (1992) were the first to propose a discrete choice task in which an individual is asked to indicate the least preferred item in a choice set, in addition to indicating the traditional most preferred one. This data collection approach is now called best-worst scaling and is applied in many fields (Scarpa *et al.*, 2011, provide several references). Items can have different nature (Flynn & Marley, 2014; Louviere *et al.*, 2015; Marley & Flynn, 2015). Initially, they were restricted to attitudes, public policy goals, brands, or anything that does not require a detailed description. Later, they were extended to more complex items, such as attributes and levels describing a single alternative, or complete profiles of multiple alternatives that are standard in choice modeling. The paper deals with this latter case.

Best and worst choices have long been analyzed using random utility models (RUM). Traditionally, distinct models have been considered for the three types of choice: best only, worst only, and best-worst. Which model should be used for best choices, which one for worst choices and which one for best-worst choices remains mainly an empirical question. However, the models used in empirical applications are often mainly driven by tractability reasons.

An i.i.d. extreme value (EV) distribution of the random terms is assumed. For best choices, the ordinary multinomial Logit (MNL) is used. For worst choices, the so-called reverse Logit model is used (see Anderson & de Palma, 1999). This model, which provides worst choice probabilities in MNL form with systematic utilities having a minus sign, originates from a form of the total utility where the random term is subtracted (instead of being added) to the systematic part. For best-worst choices, Finn & Louviere (1992) have proposed the so-called maximum difference (max-diff) model. In this model, the choice maker maximizes the difference in the utilities, with random terms specific of the alternative pair.

As elucidated by Louviere *et al.* (2015), one should estimate model forms coherent with the process used by individuals to make choices when asked. Traditionally, two such processes are considered: (i) simultaneous best and worst choice, and (ii) sequential (best-then-worst, or, less frequently, worst-then-best) choice. Thus, one should pay attention to task structures, and decide whether, for an interviewer-administered or a web-based survey, an assumption of simultaneous best and worst choices is feasible and meaningful, or the survey uses a sequential task administration process that consequently requires sequential modeling.

In practice, in the case of simultaneous best and worst choice, estimation is carried out using a maxdiff model, or by data pooling. In the latter approach, one expands the best choice dataset with worst choice data where a minus sign precedes the value of each attribute of the utilities. One estimates then a MNL with the joint likelihood. Both best only and worst only choices are modeled with reference to the complete choice set.

In the case of sequential best-then-worst choices, data pooling is common. This approach is similar to the simultaneous case, but one considers, in constructing the joint likelihood, that the worst choice is performed from a different set, because only the alternatives that remain after the best choice are considered when deciding which alternative is worst. Worst only choices are modeled with reference to this reduced choice set. The case of worst-then-best choices is analogous.

In all of the above cases, estimation is accomplished using standard software for MNL. Data pooling machinery, for both the simultaneous and the sequential case, is implemented in the most recent version of one econometric software[1].

In the paper, we restrict our attention to simultaneous best and worst choice. We consider two consistent RUMs according to the definitions given in Marley & Louviere (2005). Consistency implies that the best, worst and best-worst choices are all modeled according to a single common set of utilities. In the first model – the additive model - the random terms have a plus sign, while in the second model - the reverse model - they have a minus sign.

In best-worst scaling, there are a number of open issues. Some are theoretical. Some are empirical. The following are tackled here. On the theoretical side, results on worst and best-worst probabilities need to be extended to the multivariate normal distribution of the random terms, i.e. the Probit case. Conditions for strict log-concavity of the likelihood associated with best, worst and best-worst choice probabilities is another issue. Strictness of log-concavity is important since it ensures uniqueness of the maximum. On the empirical side, the data augmentation perspective reckons that best and worst choices have the potential to provide better estimates of the utility coefficients than best choices only. The substitutability of best and worst data, that is taken for granted in

[1] NLogit 6, by Econometric Software, Inc.

this perspective, has been called into question. The following section provides a review of the relevant literature on these issues.

1.2. Literature background

For the i.i.d. EV, i.e. Logit, distribution of the random terms, the closed-form expressions of the best, worst and best-worst choice probabilities in the additive and reverse models are known from Marley & Louviere (2005) and de Palma *et al.* (2017). In the additive model, the best choice probabilities are ordinary MNL. The worst choice probabilities are expressed as an alternating sum of best choice probabilities (also derived by Fok *et al.*, 2012, and by Vann Ophem *et al.*, 1999). The best-worst choice probabilities are a product of best choice and worst choice probability.

For the multivariate normal, i.e. Probit, distribution of the random terms, no investigation has been carried out yet on the expressions of worst and best-worst choice probabilities. Necessary and sufficient conditions for strict log-concavity of the likelihood are available, to the best of our knowledge, only for the two cases of best probabilities of the multinomial Logit (McFadden, 1974), and best probabilities of the binomial Probit (Wedderburn, 1976; Daganzo, 1979; Amemiya, 1985). For both distributions (Logit and Probit), there is no example, so far, of estimation that exhibits consistency between best, worst and best-worst choices.

The substitutability assumption between best and worst choice data has been questioned by a few authors. Louviere *et al.* (2015, Ch. 14), with examples in the marketing area, and Ben-Akiva *et al.* (1992) and Giergiczny *et al.* (2017), with examples in the transportation area, find statistically significant differences between depth-specific coefficients, i.e. coefficients based on best choice data only and coefficients based on worst choice data only. The results restrict to the Logit.

The approaches used by these authors are different. Louviere *et al.* (2015) estimate a reverse Logit model with worst choices. Ben-Akiva *et al.* (1992) and Giergiczny *et al.* (2017) use data based on full rankings, and compare the results of rank-ordered Logit models estimated with the full dataset and with conditional datasets, i.e. datasets obtained by eliminating the first-best alternative, the first- and the second-best alternative, and so on. The main result common to these studies is that the attenuation of the values of the coefficients, from best to worst choices, is a manifestation of the higher uncertainty (lower scale parameter) that affects ranking preferences of the individuals at lower depths (i.e. for worst choices). Giergiczny *et al.* (2017) provide the only analysis so far of the differences between best and worst choices in terms of value of travel time savings (VTTS). The case they address relates to rail services.

Yan & Yoo (2014) find, with simulation experiments, that the estimates produced by the rank-ordered Logit model can show coefficient attenuation if the true distribution of the random terms deviates even slightly from the i.i.d. EV distribution. This result has motivated the papers by Nair *et al.* (2018a,b), where the performances of the rank-ordered Probit are investigated. Nair *et al.*

(2018b) find the rank-ordered Probit more robust than the rank-ordered Logit in terms of stability of the coefficients across ranking depths.

1.3. Contribution

In the paper, we consider consistent additive and reverse models under two distributions of the random terms: i.i.d. EV, i.e. Logit, and multivariate normal, i.e. Probit.

In the theoretical part of the paper, we address two issues. The first is the derivation of probabilities for best, worst and best-worst choices. After a presentation of preliminary results extending to any additive RUM, we review the best, worst and best-worst choice probabilities of the two models (additive and reverse) in the Logit case. For the Probit, we provide new results for worst choice probabilities.

The second issue is the strict log-concavity, with respect to the coefficients of the systematic utilities, of the choice probabilities and of the associated likelihood functions. In particular, for the Logit, we prove strict log-concavity of the likelihood function of MNL under milder assumptions than those in McFadden (1974). For the Probit, we provide (mild) conditions for strict log-concavity of the likelihood function of multinomial Probit (MNP). Clearly, these results have a wider impact than that on best-worst scaling only.

At the same time, the paper aims to test, with a numerical example of practical relevance in transportation, the implict assumption about the substitutability of best and worst data that is taken for granted in the traditional data augmentation perspective. The test extends to Logit and Probit, and is carried out in terms of coefficients of the systematic utilities and VTTS alike.

In the empirical part, the data are presented. They relate to a stated-preference survey carried out in Rome in 2015. The choice of the transportation mode among three alternatives - car, powered-two-wheeler (PTW) and public transportation - is considered. Respondents in the sample were asked to state their best and worst choice. The survey was aimed to assess the responses of car users to charging measures, similar to those implemented in London, Stockholm and Milan. This policy is currently being planned by the municipality. Then, we present the results of the estimation of additive and reverse Logit and Probit models, each with best choice data only, worst choice data only, and best-worst data.

The remainder of the paper is organised as follows. Section 2 and 3 are theoretical, and present general assumptions and identities. Section 4 and 5 deal with specific theoretical results for Logit and Probit models, respectively. Section 6 presents the data and the estimation results. Section 7 concludes. Proofs of lemmas, propositions and theorems are relegated to appendices.

2. Main assumptions

Consider an individual facing a choice of the best and/or the worst alternative among a finite set of alternatives $X \equiv \{1, \cdots, n\}$, where $n \equiv |X|$ denotes the cardinality of X, with $n \geq 3$.

Individual preferences are represented by an additive random utility model (ARUM). We assume there exists a vector of systematic utilities $\boldsymbol{v} \equiv (v_1, \cdots, v_n)$, and a vector of random utilities $\boldsymbol{\epsilon} \equiv (\epsilon_1, \cdots, \epsilon_n)$ such that the utility of alternative z takes the following additive form

$$U_z = v_z + \epsilon_z,\ z \in X. \tag{1}$$

We assume that $\boldsymbol{\epsilon}$ admits a positive and continuous probability density function with respect to Lebesgue measure on $\mathbb{R}^n$.[2] Let $F_{\boldsymbol{\epsilon}}(\cdot)$ be the cumulative distribution function of the vector $\boldsymbol{\epsilon}$, whose arguments (realizations) are denoted by $\boldsymbol{t} \equiv (t_1, \cdots, t_n) \in \mathbb{R}^n$.

With the above assumptions, non-coincidence holds, i.e. $\mathbb{P}(U_z = U_{z'}) = 0,\ z, z' \in X,\ z' \neq z$, where $\mathbb{P}(\cdot)$ is a probability measure. In words, ties among utilities occur with zero probability.

3. Identities

3.1. Best choice probabilities

For any non-empty subset Y of X and any alternative x of Y, the probability that alternative x is the best choice in Y is given by

$$B_Y(x; \boldsymbol{v}) \equiv \mathbb{P}\left(U_x \geq U_z,\ z \in Y\right),\ x \in Y \subseteq X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{2}$$

All these probabilities are forming a *system* of best choice probabilities. The probabilities in the whole set can be obtained by performing the following single integration (S.P. *et al.*, 1992, Eq 2.27)

$$B_X(x; \boldsymbol{v}) = \int_{-\infty}^{+\infty} \frac{\partial F_{\boldsymbol{\epsilon}}(u\mathbf{1}_X - \boldsymbol{v})}{\partial t_x} du,\ x \in X,\ \boldsymbol{v} \in \mathbb{R}^n, \tag{3}$$

where $\mathbf{1}_X$ is the all-ones n-vector. The best choice probabilities on any set $Y \subseteq X$ can be obtained by performing the following limit[3]

$$B_Y(x; \boldsymbol{v}) = \lim_{\substack{v_z \to -\infty \\ \forall z \in X \setminus Y}} B_X(x; \boldsymbol{v}),\ x \in Y \subseteq X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{4}$$

3.2. Worst choice probabilities

Consider now the system of worst choice probabilities defined by

$$W_Y(x; \boldsymbol{v}) \equiv \mathbb{P}\left(U_z \geq U_x,\ z \in Y\right),\ x \in Y \subseteq X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{5}$$

[2] That is, the support of $\boldsymbol{\epsilon}$ is the whole space $\mathbb{R}^n$.

[3] A proof of this equality requires the use of Fubini's theorem to interchange the order of integrals and Lebesgue's dominated convergence theorem to interchange the limit with an integral sign.

A general identity provided in de Palma *et al.* (2017, Theorem 5) allows to recover the worst choice probabilities from the best choice probabilities. Accordingly, worst choice probabilities are given by an alternating sum of best choice probabilities

$$W_X(x;\boldsymbol{v}) = \sum_{\{x\}\subseteq Y\subseteq X} (-1)^{|Y|-1} B_Y(x;\boldsymbol{v}),\ x \in X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{6}$$

3.3. Best-worst choice probabilities

Now, for two different alternatives x and y of X, consider the probability that, jointly, x is the best in X and alternative y is the worst in X. Formally, best-worst choice probabilities are defined by

$$BW_X\left(x,y;\boldsymbol{v}\right) \equiv \mathbb{P}\left(U_x \geq U_z \geq U_y,\ z \in X\right),\ x,y \in X,\ x \neq y,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{7}$$

Let $\mathbf{1}_Y$ denotes the n-vector with zth component being one if $z \in Y$ and zero if $z \in X \setminus Y$, with $Y \subseteq X$. We have established this new identity:

Lemma 1. *The best-worst choice probabilities are, for any $x, y \in X$, $x \neq y$, and any $\boldsymbol{v} \in \mathbb{R}^n$, given by:*

$$BW_X\left(x,y;\boldsymbol{v}\right) = \sum_{\{y\}\subseteq Y\subseteq X\setminus\{x\}} (-1)^{|Y|-1} B_{X,Y}(x,y;\boldsymbol{v}), \tag{8}$$

where:

$$B_{X,Y}(x,y;\boldsymbol{v}) \equiv \int_0^\infty \frac{-\partial B_X(x;\boldsymbol{v}+t\mathbf{1}_Y)}{\partial v_y} dt. \tag{9}$$

Proof. See Appendix A.

$B_{X,Y}(x,y;\boldsymbol{v})$ represents the probability that, jointly, x is the best alternative in X and y is the best alternative in Y (see de Palma *et al.*, 2017).

3.4. Log-likelihood identities and properties

We consider the case where the systematic part of the utility of alternative z has the following linear specification

$$v_z = \boldsymbol{\xi}_z \boldsymbol{\theta},\ z \in X, \tag{10}$$

where $\boldsymbol{\xi}_z$ is a row vector of alternative and individual characteristics of $\mathbb{R}^K$, and $\boldsymbol{\theta}$ is a column vector of coefficients of $\mathbb{R}^K$ to be estimated.

The log-likelihood associated with the observation of x as the best (resp. worst) choice in Y is $\mathcal{L}_Y^B(x) \equiv \ln B_Y(x;\boldsymbol{v})$ (resp. $\mathcal{L}_Y^W(x) \equiv \ln W_Y(x;\boldsymbol{v})$), $x \in Y \subseteq X$. In the sequel, $\nabla_{\boldsymbol{\theta}}\mathcal{L}_Y^B(x)$ (resp. $\nabla_{\boldsymbol{\theta}}\mathcal{L}_Y^W(x)$) denotes the row gradient of $\mathcal{L}_Y^B(x)$ (resp. $\mathcal{L}_Y^W(x)$) with respect to $\boldsymbol{\theta}$. Moreover, we denote by $\nabla_{\boldsymbol{\theta}}^2\mathcal{L}_Y^B(x)$ and $\nabla_{\boldsymbol{\theta}}^2\mathcal{L}_X^W(x)$ the corresponding Hessian matrices. The vector argument $\boldsymbol{v}$ is omitted for the sake of notational convenience. Evaluating the gradient vector and the Hessian matrix of the log-likelihood functions is critical for estimation purpose.

For any row vector $\boldsymbol{\xi}$ of $\mathbb{R}^K$, we shall use the vectorial notation $\boldsymbol{\xi}^2 \equiv \boldsymbol{\xi}'\boldsymbol{\xi}$, where the prime denotes vector transpose. Using the following alternating weights[4]

$$\omega_{X,Y}(x) \equiv \frac{(-1)^{|Y|-1} B_Y(x;\boldsymbol{v})}{W_X(x;\boldsymbol{v})},\ x \in Y \subseteq X, \tag{11}$$

we establish a result which provide identities relating to the gradient vectors and the Hessian matrix of the log-likelihood functions associated with best and worst choices.

Lemma 2. *For any ARUM with linear specification (10), for any $x \in X$, we have:*

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \sum_{\{x\}\subseteq Y \subseteq X} \omega_{X,Y}(x) \nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x). \tag{12}$$

Moreover,

$$\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) + \left[\nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x)\right]^2 = \sum_{\{x\}\subseteq Y \subseteq X} \omega_{X,Y}(x) \left\{\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) + \left[\nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x)\right]^2\right\}. \tag{13}$$

Proof. See Appendix A.

This lemma shows that, for general ARUMs, the gradient of the log-likelihood associated with worst choice is an alternating weighted sum of gradients of the log-likelihood associated with best choice. The Hessian associated with worst choice involves a similar alternating weighted sum of best choice Hessians plus square matrices obtained from differences between best and worst choice gradients.

3.5. Reverse models

Now, we introduce a parallel RUM where the roles of best and worst choice probabilities are interchanged. Following Anderson and de Palma (1999), we associate to every RUM its corresponding reverse ARUM, where the utilities are now given by

$$U_z^R = v_z - \epsilon_z,\ z \in X. \tag{14}$$

Note the term $-\epsilon_z$, whereas in the previous model (see Eq. (1)) we had $+\epsilon_z$.

Clearly, the best choice best choice probabilities, in the reverse model, defined by $B_X^R(x,\boldsymbol{v}) \equiv \mathbb{P}\left(U_x^R \geq U_z^R,\ z \in X\right)$, satisfy the following

$$B_X^R(x,\boldsymbol{v}) = W_X(x,-\boldsymbol{v}),\ x \in X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{15}$$

[4] The vector argument $\boldsymbol{v}$ is voluntarily omitted from the notation of the alternating weights. Thanks to Identity (6), we can verify that: $\sum_{\{x\}\subseteq Y \subseteq X} \omega_{X,Y}(x) = 1,\ x \in X$.

Conversely, the worst choice probabilities in the reverse model, defined by $W_X^R(x,\boldsymbol{v}) \equiv \mathbb{P}\left(U_z^R \geq U_x^R,\ z \in X\right)$, satisfy

$$W_X^R(x,\boldsymbol{v}) = B_X(x,-\boldsymbol{v}),\ x \in X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{16}$$

Finally, the best-worst choice probabilities in the reverse model, defined by $BW_X^R(x,y;\boldsymbol{v}) \equiv \mathbb{P}\left(U_x^R \geq U_z^R \geq U_y^R,\ z \in X\right)$, are related to the best-worst choice probabilities in the original model by

$$BW_X^R(x,y;\boldsymbol{v}) = BW_X(y,x;-\boldsymbol{v}),\ x,y \in X,\ x \neq y,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{17}$$

4. Logit

4.1. Best, worst and best-worst choice probabilities

Assume the random variables ϵ_x, $x \in X$, follows independent standard Gumbel distributions. That is, the cumulative distribution function of $\boldsymbol{\epsilon}$ is given by

$$F_{\boldsymbol{\epsilon}}(\boldsymbol{t}) = \exp(-\sum_{z \in X} e^{-t_z}),\ \boldsymbol{t} \in \mathbb{R}^n. \tag{18}$$

Differentiating (18), then applying (3) and (4), best choice probabilities are obtained by integration yielding the ordinary multinomial Logit model (MNL) formula[5]

$$B_Y\left(x;\boldsymbol{v}\right) = \frac{e^{v_x}}{\sum_{z \in Y} e^{v_z}},\ x \in Y \subseteq X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{19}$$

It is worth to rewrite the above Logit expression as follows

$$B_Y\left(x;\boldsymbol{v}\right) = e^{v_x - \mathit{\Lambda}_Y(\boldsymbol{v})},\ x \in Y \subseteq X,\ \boldsymbol{v} \in \mathbb{R}^n, \tag{20}$$

where $\mathit{\Lambda}_Y(\boldsymbol{v}) \equiv \ln \sum_{z \in Y} e^{v_z}$ is the Log-Sum-Exp function. Note the following property (Williams-Daly-Zachary theorem; McFadden, 1981)

$$\frac{\partial \mathit{\Lambda}_Y(\boldsymbol{v})}{\partial v_z} = \begin{cases} B_Y\left(z;\boldsymbol{v}\right), & z \in Y \subseteq X; \\ 0, & z \in X \setminus Y. \end{cases} \tag{21}$$

Applying Identity (6), the corresponding worst choice probabilities are obtained (also found in Fok *et al.*, 2012, and in Vann Ophem *et al.*, 1999)

$$W_X\left(x;\boldsymbol{v}\right) = \sum_{\{x\} \subseteq Y \subseteq X} (-1)^{|Y|-1} \frac{e^{v_x}}{\sum_{z \in Y} e^{v_z}},\ x \in X,\ \boldsymbol{v} \in \mathbb{R}^n. \tag{22}$$

[5]The integrand of the RHS of (3) is given by: $\partial F_{\boldsymbol{\epsilon}}(u\mathbf{1}_X - \boldsymbol{v})/\partial t_x = e^{v_x - u}\exp(-e^{\mathit{\Lambda}_X(\boldsymbol{v}) - u})$. By integration, we get: $B_X(x,\boldsymbol{v}) = e^{v_x - \mathit{\Lambda}_X(\boldsymbol{v})}$. The remaining MNL best choice probabilities are then obtained by applying (4).

Note that the alternating weights defined by (11) are explicitly given by

$$\omega_{X,Y}(x) \equiv \frac{(-1)^{|Y|-1} e^{-\Lambda_Y(\boldsymbol{v})}}{\sum_{\{x\}\subseteq Z\subseteq X} (-1)^{|Z|-1} e^{-\Lambda_Z(\boldsymbol{v})}},\ x \in Y \subseteq X. \tag{23}$$

Best-worst choice probabilities are, for any $x, y \in X$, $x \neq y$, and any $\boldsymbol{v} \in \mathbb{R}^n$, given by[6]

$$BW_X(x, y; \boldsymbol{v}) = \frac{e^{v_x}}{\sum_{z\in X} e^{v_z}} \sum_{\{y\}\subseteq Y\subseteq X\setminus\{x\}} (-1)^{|Y|-1} \frac{e^{v_y}}{\sum_{z\in Y} e^{v_z}}. \tag{24}$$

When we have three alternatives only, it is easily shown that the additive Logit with B&W choices is exactly the same as the rank-ordered Logit. In fact, with reference to alternatives 1 and 2, we have:

$$\begin{aligned} BW_{\{1,2,3\}}(1,2) &= \frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}} \left(1 - \frac{e^{v_2}}{e^{v_2}+e^{v_3}}\right) \\ &= \frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}} \frac{e^{v_3}}{e^{v_2}+e^{v_3}} \\ &= B_{\{1,2,3\}}(1)\, B_{\{2,3\}}(3), \end{aligned} \tag{25}$$

which is the rank-ordered Logit model, where best choice probabilities are exploded.

Our results can be put into perspective with the impossibility theorem by Luce & Suppes (1965). The theorem assumptions are that: (i) ranking probabilities are derived from a common set of utilities, (ii) the probability of the ranking from best to worst is obtained from explosion of best choice probabilities, and the probability of the ranking from worst to best is also obtained from explosion of worst choice probabilities. The theorem states that, under these assumptions, best choice probabilities should be equal across alternatives, and worst choice probabilities should also be equal across alternatives, an impossibility result. Notice that not all the assumptions of this theorem apply to the framework here. In the additive Logit model, explosion of best probabilities, for computation of the probability of the ranking from best to worst, holds because best probabilities are MNL, but explosion of worst probabilities, for computation of the probability of the ranking from worst to best, does not because they are not MNL. With three alternatives, we do not compute the probability of the ranking from worst to best by exploding worst probabilities, but we use the formula of best-worst choice probability.

[6] Since $\partial B_X(x; \boldsymbol{v} + t\mathbf{1}_Y)/\partial v_y = -e^{v_x+v_y+t}/[e^{\Lambda_{X\setminus Y}(\boldsymbol{v})} + e^{t+\Lambda_Y(\boldsymbol{v})}]^2$, integration of that expression yields (notice that $e^{\Lambda_{X\setminus Y}(\boldsymbol{v})} + e^{\Lambda_Y(\boldsymbol{v})} = e^{\Lambda_X(\boldsymbol{v})}$): $B_{X,Y}(x, y; \boldsymbol{v}) = e^{v_x+v_y-\Lambda_X(\boldsymbol{v})-\Lambda_Y(\boldsymbol{v})}$. Applying then Lemma 1 yields the expression of the MNL best-worst choice probabilities.

4.2. Log-likelihood

Let us recall a classical result in the econometrics of the MNL (see, e.g., McFadden, 1974; Ben-Akiva & Lerman, 1985). Note that the log-likelihood corresponding to the observation of the best choice is, using (20), given by

$$\mathcal{L}_Y^B(x) = v_x - \Lambda_Y(\boldsymbol{v}),\ x \in Y \subseteq X.$$

Therefore, using (21), its gradient satisfies

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) = \boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_Y,\ x \in Y \subseteq X, \tag{26}$$

where

$$\overline{\boldsymbol{\xi}}_Y \equiv \sum_{z \in Y} B_Y\left(z; \boldsymbol{v}\right) \boldsymbol{\xi}_z,\ Y \subseteq X. \tag{27}$$

The corresponding Hessian matrix is given by

$$\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) = H_Y(\boldsymbol{v}) \equiv - \sum_{z \in Y} B_Y\left(z; \boldsymbol{v}\right) (\boldsymbol{\xi}_z - \overline{\boldsymbol{\xi}}_Y)^2,\ Y \subseteq X. \tag{28}$$

It is denoted by $H_Y(\boldsymbol{v})$ since it does not depend on the alternative x.

From (28), the Hessian matrix of the best choice log-likelihood function is a semi-definite negative matrix, because it is, formally, the negative of a covariance matrix.

For the worst choice, the formulas are more intricate. We shall use the following notation

$$\overline{\boldsymbol{\xi}}_X^W(x) \equiv \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \overline{\boldsymbol{\xi}}_Y,\ x \in X. \tag{29}$$

Note that an alternative form of $\overline{\boldsymbol{\xi}}_X^W(x)$ is

$$\overline{\boldsymbol{\xi}}_X^W(x) = \sum_{z \in X} \theta_X(x, z; \boldsymbol{v}) \boldsymbol{\xi}_z,\ x \in X, \tag{30}$$

where

$$\theta_X(x, z; \boldsymbol{v}) \equiv \sum_{\{x,z\} \subseteq Y \subseteq X} (-1)^{|Y|-1} \frac{B_Y(x; \boldsymbol{v}) B_Y(z; \boldsymbol{v})}{W_X(x; \boldsymbol{v})},\ x, z \in X. \tag{31}$$

We have:

Proposition 1. *For a MNL with linear specification (10), for any $x \in X$, we have:*

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_X^W(x). \tag{32}$$

Moreover, the corresponding Hessian matrix is:

$$\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left\{ H_Y(\boldsymbol{v}) + \left[\overline{\boldsymbol{\xi}}_Y - \overline{\boldsymbol{\xi}}_X^W(x) \right]^2 \right\}. \tag{33}$$

Proof. See Appendix B.

It is far from being clear from (33) whether the Hessian matrix is semi-definite. At this stage, we will go even further and establish Proposition 2 which will be helpful to find some mild conditions ensuring that the log-likelihood functions are strictly concave (Theorem 1). Strict concavity ensures uniqueness of the maximum.

We denote by $\boldsymbol{d} \equiv (d_1, \cdots, d_{n-1}) \in \mathbb{R}^{n-1}$ the vector whose components are $d_z = v_z - v_n$, $z = 1, \cdots n-1$. We are using a Prékopa's theorem (Prékopa, 1973) which states that if the probability density function $f(\cdot)$ is positive and strictly logarithmic concave function in an open convex set of $\mathbb{R}^{n-1}$, then the associated cumulative distribution function $F(\cdot)$ is also strictly logarithmic concave in the same set. We have:

Proposition 2. *The MNL best choice probabilities $B_X(x; \boldsymbol{v})$, given by the Logit formula (19), and the worst choice probabilities $W_X(x; \boldsymbol{v})$ given by (22), are strictly log-concave in $\boldsymbol{d}$, for any $x \in X$.*

Proof. See Appendix B.

We consider now a sample of N observations indexed by $i = 1, \cdots, N$. The observation index will be in general omitted in the notations below. Each observation consists of a best and a worst choice within the given choice set X (not necessarily the same for all the observations in the sample). Every observation provides log-likelihood functions, $\mathcal{L}^B{}_X(x)$ and $\mathcal{L}^W{}_X(y)$, where $x, y \in X$ are the observed best and worst choices, respectively. They contribute to the total best and worst log-likelihood functions, $\mathcal{L}^B$ and $\mathcal{L}^W$, which are obtained by summing-up, over the whole sample, those terms.

We denote by $\mathcal{M}$ the matrix obtained by concatenating the rows of the N observation-specific $\left[(n^i - 1) \times K\right]$ matrices whose rows are $\boldsymbol{\xi}_z - \boldsymbol{\xi}_{n^i}$, $z \in X^i \setminus \{n^i\}$, where X^i denotes the choice set of i and n^i its cardinality. The matrix $\mathcal{M}$ has $\sum_{i=1}^N n^i - N$ rows and K columns.

Theorem 1. *In the linear-in-the-coefficients MNL model, the total log-likelihood functions of the best choices $\mathcal{L}^{\mathcal{B}}$ and of the worst choices $\mathcal{L}^{\mathcal{W}}$ are strictly concave in $\boldsymbol{\theta}$ iff the matrix $\mathcal{M}$ has full rank.*

Proof. See Appendix B.

Note the high simplicity of these necessary and sufficient conditions for strict log-concavity of both $\mathcal{L}^B$ and $\mathcal{L}^W$. Recall that the conditions in McFadden (1974, Axiom 5) for the global concavity of $\mathcal{L}^B$ require the full rank of the $\sum_{i=1}^N n^i \times K$ matrix whose rows are $\boldsymbol{\xi}_z - \overline{\boldsymbol{\xi}}_{X^i}$. Since $\overline{\boldsymbol{\xi}}_{X^i}$ depends on $\boldsymbol{\theta}$, McFadden's conditions are less easy to check.

We have therefore proved that the likelihood functions associated with best and worst choices have, under the stated conditions, a unique maximum, provided one exists. Similar results can be derived for best-worst choices, on the basis of (24), by preservation of strict log-concavity under product.

4.3. Reverse Logit

Consider now the reverse MNL. Applying (15) and (22), the best choice probabilities are given by (see also the derivation of Anderson & de Palma, 1999):

$$B_X^R(x;\boldsymbol{v}) = \sum_{\{x\}\subseteq Y\subseteq X} (-1)^{|Y|-1} \frac{e^{-v_x}}{\sum_{z\in Y} e^{-v_z}},\ x \in X. \tag{34}$$

Likewise, the worst choice probabilities are given by the ordinary MNL formula with systematic utilities preceded by a minus sign (Anderson & de Palma, 1999; Marley & Louviere, 2005):

$$W_X^R(x;\boldsymbol{v}) = \frac{e^{-x}}{\sum_{z\in X} e^{-v_z}},\ x \in X. \tag{35}$$

Best-worst choice probabilities are, for $x, y \in X$, $x \neq y$, and on the basis of (17) and (24), given by (see also Marley & Louviere, 2005):

$$BW_X^R(x,y;\boldsymbol{v}) = \frac{e^{-v_y}}{\sum_{z\in X} e^{-v_z}} \sum_{\{x\}\subseteq Y\subseteq X\setminus\{y\}} (-1)^{|Y|-1} \frac{e^{-v_x}}{\sum_{z\in Y} e^{-v_z}}. \tag{36}$$

The results for the log-likelihood function are specular and are, for brevity, omitted.

5. Probit

5.1. Best, worst and best-worst choice probabilities

Assume that the vector of random components $\boldsymbol{\epsilon}$ has a multivariate normal (MVN) distribution with zero mean and non-singular covariance matrix $\mathbf{\Sigma}$. This yields the multinomial Probit (MNP) model. If the covariance matrix $\mathbf{\Sigma}$ were singular, the distribution of the random components would be degenerate, and would not admit a probability density function with respect to Lebesgue measure, contravening our initial assumption.

In the case of MNP, worst choice probabilities can be directly obtained from best choice probabilities.

Proposition 3. *In the MNP, we have:*

$$W_X(x,\mathbf{v}) = B_X(x,-\mathbf{v}),\ \mathbf{v} \in \mathbb{R}^n, x \in X \tag{37}$$

Proof. See Appendix B. □

In words, the worst choice probabilities are equal to the best choice probabilities computed with systematic utilities preceded by a minus sign.

When we have three alternatives only, the Probit with best-worst choices is the ordinary rank-ordered Probit.

5.2. Log-likelihood

In the literature, we find the following result: in the binomial Probit, the total likelihood is strictly log-concave in the estimation coefficients (see Wedderburn, 1976; Daganzo, 1979; Amemiya, 1985). For this result to hold, it suffices to assume that the matrix of observations has column full rank, otherwise there would be a non-uniqueness. No result is found relating to the strict log-concavity of total likelihood of MNP. We are able to prove the strict log-concavity, with respect to the coefficients of the systematic utilities, of the total likelihood of MNP under the same assumptions as those for strict log-concavity of total likelihood of MNL.

Proposition 4. *The MNP best choice probabilities* $B_X(x, \mathbf{v})$ *and worst choice probabilities* $W_X(x, \mathbf{v})$ *are strictly log-concave in* $\mathbf{d}$, *for any* $x \in X$.

Proof. See Appendix B.

As in the Logit, recall the $\mathcal{M}$ matrix obtained by concatenating the rows of the N observation-specific $(n^i - 1) \times K$ matrices whose rows are $\boldsymbol{\xi}_z - \boldsymbol{\xi}_{n^i}$, $z \in X^i \setminus \{n^i\}$. It plays a key role to obtain well behaved likelihood functions.

Theorem 2. *In the linear-in-the-coefficients MNP model, the total log-likelihood functions of the best choices* $\mathcal{L}^{\mathcal{B}}$ *and of the worst choices* $\mathcal{L}^{\mathcal{W}}$ *are strictly concave in* $\boldsymbol{\theta}$ *iff the matrix* $\mathcal{M}$ *has full rank.*

Proof. See Appendix B.

5.3. Reverse Probit

Consider now the reverse MNP. The following equalities hold between reverse and additive choice probabilities.

Proposition 5. *In the MNP, we have:*

$$B_X^R(x, \mathbf{v}) = B_X(x, \mathbf{v}), \ \mathbf{v} \in \mathbb{R}^n, x \in X, \tag{38}$$

and

$$W_X^R(x, \mathbf{v}) = W_X(x, \mathbf{v}), \ \mathbf{v} \in \mathbb{R}^n, x \in X. \tag{39}$$

Proof. See Appendix B. □

6. Empirical illustration

6.1. Data

In Rome, an access charging policy for a large central area, delimited by the railway ring (“Anello ferroviario”), with an extension of 33.7 square km, has been included in the new urban traffic master plan (“Piano Generale del Traffico Urbano” - PGTU) passed in 2015. The planned policy, not yet put into action, is similar to the congestion charge implemented since 2003 in London (ca 21 square km), the charging scheme implemented since 2006 in central Stockholm (ca 30 square km), and the so-called “area C” implemented since 2012 in central Milan (ca 8 square km), since it is based on a daily access charge to enter the inner city area.

A combined revealed-preference (RP) and stated preference (SP) survey was conducted, by means of computer-assisted face-to-face interviews, in Spring 2015. The sample is drawn from a population including individuals who travel by car, at least once a week, from an origin located in the area between the road ring (“Grande raccordo anulare”, GRA) and the railway ring, to a destination located inside the railway ring. The RP part investigates current travel habits and preferences for transportation modes.

The SP part considers hypothetical scenarios characterised by access charges for car and powered two wheelers (PTW), the improvement of bus service supply, and the ban of more polluting vehicles in the charged area. The choice experiment proposes three alternatives (car, PTW, public transportation) with the attributes and levels reported in Table 1.

In Rome, a significant shift towards PTW is expected from policies that create negative incentives to the use of the car. Therefore, the charging policy targets PTW as well, because shifting demand towards public transportation is a main policy goal of the municipality. To increase the realism of the interview, the travel time used for car and PTW is the perceived value stated in the RP part of the interview. For public transportation, the attribute travel time is pivoted around the revealed value with variations of -20% and -40%.

Table 1: Attributes and levels in the choice experiment

Attribute	Alt.		
	car	PTW	public transportation
access charge (EUR/day)	0,2,4,6	0,2,4,6	-
travel time (minutes)	RP	RP	RP, -20% RP, -40% RP

The experimental design is a full factorial of 48 combinations. Four combinations are excluded given the aim of the survey (dominated alternatives). After a first pilot, choice combinations were administered, for practical purposes linked to a maximum time span that had to be guaranteed in order for interviewees to complete the questionnaire, using a blocking strategy. The entire design has been subdivided into eleven blocks each including four choice tasks. This made the administration of the interview possible also in logistically unfavorable conditions, e.g. on street, at office entrance/exit, in shopping malls during regular opening hours. Figure 1 shows a choice task example. The interviewee is invited to select, among the three proposed alternatives, the two that she/he would choose with highest probability (best alternative) and lowest probability (worst alternative).

We acquired 146 interviews, each including four choice tasks. The minimum age of the interviewees is 23 years, maximum 79, mean 46.2 and standard deviation 11.3. Table 2 reports sample statistics for other socio-economic characteristics.

attributes\alternatives	car	PTW	public transportation
access charge (EUR/day)	4	2	-
travel time (minutes)	30	20	45
best alternative	○	○	○
worst alternative	○	○	○

Figure 1: Choice task example

Table 2: Sample statistics

	Abs. freq.	Percent
Gender		
male	65	44.5
female	81	55.5
Education		
primary school diploma	1	0.7
middle school diploma	4	2.7
high school diploma	60	41.1
university degree	51	34.9
higher degree (master, PhD,...)	30	20.5
Occupation		
self-employed, manager	43	29.4
employee	91	62.3
specialized worker, shopkeeper, craftsman	6	4.1
worker, retired, occasional job	2	1.3
housewife, student, unemployed	4	2.7
Net income (EUR/month)		
$\leq$500	3	2.0
501-1000	15	10.2
1001-1500	59	40.4
1501-2500	52	35.6
2501-5000	15	10.2
>5000	2	1.4
Number of respondents	146	100.0

6.2. Estimation

Table 3 reports estimation results for the three models: additive Logit, reverse Logit and Probit, with B only, W only and B&W choice data. Recall that additive Probit is identical to reverse Probit. Probit also considers independent random terms across alternatives. Alternative 1 is car, alternative 2 is PTW, alternative 3 is public transportation. The systematic utilities have the standard linear-in-the-coefficients structure.

Table 3: Estimation results: coefficients

		B		W		B&W	
	Alt.	Coefficient	(t-stat.)	Coefficient	(t-stat)	Coefficient	(t-stat)
Additive Logit							
charge (EUR/day)	1,2	-.947***	(-7.85)	-.375***	(-5.32)	-.505***	(-7.75)
travel time (minutes)	1,2,3	-.095***	(-9.82)	-.060***	(-9.74)	-.070***	(-12.55)
alt. specific constant	2	-2.186***	(-11.06)	-1.029***	(-8.21)	-1.282***	(-10.93)
alt. specific constant	3	-.453***	(-2.69)	.220	(1.36)	.094	(.78)
travel time × gender (female = 1)	1,2,3	.043***	(4.81)	.017***	(3.23)	.023***	(4.49)
charge × income (six income classes: 1,...,6)	1,2	.121***	(4.20)	.035*	(1.90)	.053***	(3.12)
583 (pseudo)observations, log-likelihood function =		-421.76		-501.35		-825.77	
Reverse Logit							
charge (EUR/day)	1,2	-.796***	(-7.69)	-.432***	(-5.27)	-.576***	(-8.08)
travel time (minutes)	1,2,3	-.081***	(-9.80)	-.072***	(-9.92)	-.071***	(-11.70)
alt. specific constant	2	-1.778***	(-11.03)	-1.164***	(-8.31)	-1.317***	(-11.15)
alt. specific constant	3	-.333**	(-2.20)	.384*	(1.86)	-.028	(-.22)
travel time × gender (female = 1)	1,2,3	.037***	(5.06)	.023***	(3.60)	.029***	(5.29)
charge × income (six income classes: 1,...,6)	1,2	.103***	(4.17)	.040*	(1.83)	.065***	(3.62)
583 (pseudo)observations, log-likelihood function =		-422.74		-503.42		-804.78	
Probit							
charge (EUR/day)	1,2	-.693***	(-8.21)	-.328***	(-5.32)	-.443***	(-6.69)
travel time (minutes)	1,2,3	-.073***	(-10.24)	-.055***	(-10.39)	-.060***	(-11.89)
alt. specific constant	2	-1.644***	(-11.56)	.940***	(8.55)	-1.129***	(-12.03)
alt. specific constant	3	-.323**	(-2.39)	-.229	(-1.48)	0.025	(0.24)
travel time × gender (female = 1)	1,2,3	.033***	(5.03)	.016***	(3.38)	0.022***	(4.54)
charge × income (six income classes: 1,...,6)	1,2	.085***	(4.16)	.027*	(1.67)	.045***	(2.57)
583 (pseudo)observations, log-likelihood function =		-421.11		-501.59		-811.75	
Key: ***,**,* = statistical significance at 1%, 5%, 10% level							

Remember that, since we have three alternatives only, the additive Logit with B&W choices is exactly the same as the rank-ordered Logit. Similarly, the Probit with B&W choices is the ordinary rank-ordered Probit.

In all three models with all choice data we observe the following common traits.

- The congestion charge and travel time coefficients have the expected sign and are statistically significant (1% level).

- The alternative specific constant of alternative 2 (PTW) is negative and statistically significant (1% level) (with the exception of Probit with W choice data which shows a sign reversal), the reference alternative being fixed as alternative 1 (car); the negative sign can be explained with the lower levels of comfort and safety.

- The coefficient associated with the travel time-gender interaction effect is positive and statistically significant (1% level), denoting an attitude of females to accept longer travel times than males.

- The coefficient associated with the charge-income interaction effect is positive, denoting that, as expected, income increases reduce the disutility of the congestion charge.

The first analysis relates to the stability of the coefficients across B only and W only choice data.

The magnitude of most coefficients is affected by the choice data, something which would have significant impacts on best choice shares, if data aggregation were post-performed upon estimation results. The estimates with W choices appear attenuated in comparison with the B choice ones. The statistical significance, measured by the absolute value of the t-statistic, of the coefficients with W data is lower than with B data. In this regard, one notes that scale differences can exist between B and W data (Swait & Louviere, 1993). This makes direct comparison between coefficients, and the associated statistical significance, not possible.

Figure 2, 3 and 4 show for the additive Logit, reverse Logit and Probit respectively, the plot of the W choice versus B choice coefficients. For the additive Logit, the linear regression is $worst = .472 \cdot best + .078$, with $R^2 = .84$; for the reverse Logit, the linear regression is $worst = .667 \cdot best + .113$, with $R^2 = .79$; for the Probit the linear regression is not estimated due to a sign reversal in one coefficient. The slope of the best-fit linear regression line is less than unity, another manifestation of the attenuation in the coefficients. It is as if, when moving from B to W, the random term variance increased, because individuals are less certain about their least preferred alternative. This outcome is an empirical regularity in rank-ordered data analysis (see the results in Ben-Akiva *et al.*, 1992, Louviere *et al.*, 2015, Ch. 14, and Giergiczny *et al.*, 2017), which is explained with the relative cognitive difficulties of identifying best and worst alternatives.

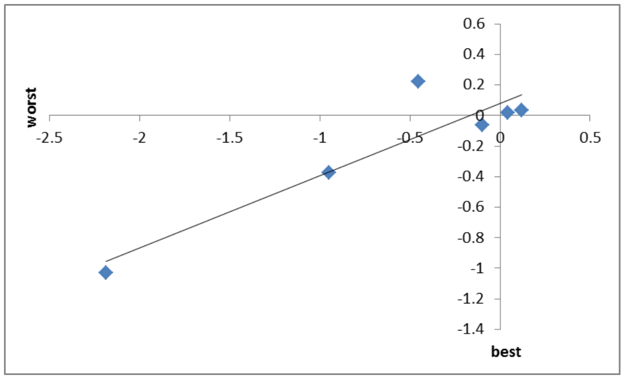

Figure 2: W choice coefficients plotted against B choice coefficients: additive Logit

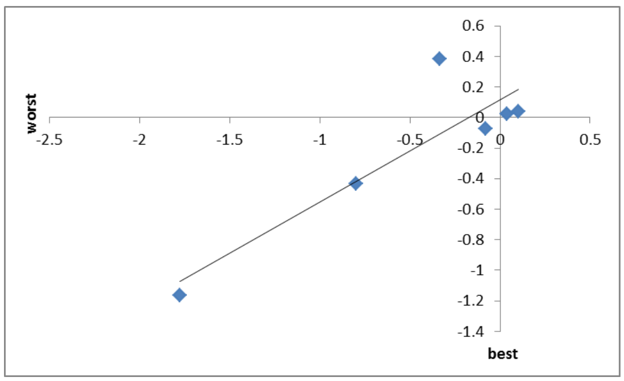

Figure 3: W choice coefficients plotted against B choice coefficients: reverse Logit

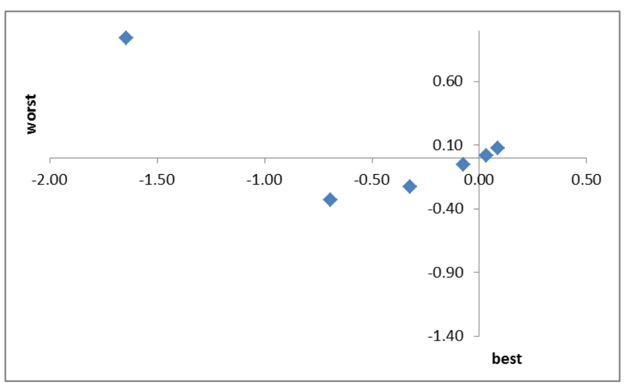

Figure 4: W choice coefficients plotted against B choice coefficients: Probit

The equality of individual coefficients between B choice data and W choice data is tested by the asymptotically normal test statistic:

$$\frac{\beta_{kB} - \beta_{kW}}{\sqrt{Var(\beta_{kB}) + Var(\beta_{kW})}} \tag{40}$$

The results of this test are in Table 4. Most differences are significant at 5% level. Statistical significance of differences results to be lower in Probit than in additive Logit (with the exception of one coefficient), a result in agreement with Nair *et al.* (2018b), who find the rank-ordered Probit a superior option to the rank-ordered Logit in terms of robustness of coefficients across ranking depths.

Table 4: Asymptotically normal test statistic: B-W difference of coefficients

	Alt.	Additive Logit	Reverse Logit	Probit
charge	1,2	-4.09***	-2.76***	-3.50***
travel time	1,2,3	-3.02***	-0.75	-1.99**
alt. specific constant	2	-4.94***	-2.87***	-14.38***
alt. specific constant	3	-2.88***	-2.80***	-0.46
travel time × gender	1,2,3	2.43**	1.47	2.02**
charge × income	1,2	2.50**	1.91*	2.18**
Key: ***,**,* = statistical significance at 1%, 5%, 10% level				

We have conducted additional simulations to assess whether the coefficient attenuation is attributable to the model's mathematical characteristics or to the individuals' behaviour. To this aim, a vector of coefficients of the systematic utilities is assumed. Then the random terms are drawn and the B and the W alternative identified. Two models are estimated with these simulated choices: one with B choices and one with W choices. Three cases are considered: additive Logit, reverse Logit and Probit. As a result, no systematic bias, i.e. no coefficient attenuation from B to W, is found with simulated choices. We can infer this empirical finding to be a behavioural effect, rather than a mathematical artefact.

A second analysis relates to the values of time. The values of the marginal rate of substitution between travel time and the congestion charge (from the utilities of the car and PTW alternatives) can be assumed, in the context of the example here, as a measure of the marginal willingness to pay (WTP). This WTP expresses the value of the travel time savings (VTTS). WTP values, and therefore VTTS, are not affected by the scale value.

The mean VTTS are shown in Table 5 for the additive Logit, reverse Logit and Probit. In all three models, mean VTTS with W choices are higher than mean VTTS with B choices (except for the highest male income class). In contrast, Giergiczny *et al.* (2017) find for the choice of rail services database the opposite result.

Approximated standard errors of the VTTS are also computed, using the delta method (Greene, 2012). The associated t-statistics are shown in Table 5. The VTTS confidence intervals with W data are relatively larger than those with B data, except for the highest income class.

Alternative methods can be used to compute WTP confidence intervals. Gatta *et al.* (2015) provide a systematic comparison of eleven methods investigating the performance under different conditions. They suggest using the likelihood ratio test inversion method, which is not affected by the cost coefficient approaching zero and is robust to small departures from correct specification, especially with small samples. The delta method, however, produces similar results when relatively large samples are involved, as in the case under consideration.

The equality of VTTS values between B choice data and W choice data is tested by the asymptotically normal test statistic. Table 6 reports the results of the test. Statistically significant differences between B and W choice data tend to be shown by lower income classes and the female population. Statistical significance of differences results to be lower in Probit than in additive Logit, a result reinforcing the agreement with Nair *et al.* (2018b), who find, as observed, the rank-ordered Probit a superior option to the rank-ordered Logit in terms of robustness of coefficients across ranking depths.

Table 5: Estimation results: VTTS (EUR/h)

	Additive Logit				Reverse Logit				Probit			
Net income	B		W		B		W		B		W	
(EUR/month)	Mean	(t-stat.)	Mean	(t-stat.)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Male												
$\leq$ 500	6.9	(6.74)	10.7	(5.52)	7.0	(6.58)	11.1	(5.44)	7.28	(6.85)	11.17	(5.51)
501-1000	8.1	(7.46)	11.9	(6.54)	8.2	(7.33)	12.4	(6.46)	8.46	(7.58)	12.30	(6.54)
1001-1500	9.8	(8.30)	13.5	(7.51)	10.0	(8.30)	14.0	(7.43)	10.11	(8.43)	13.69	(7.47)
1501-2500	12.4	(8.35)	15.6	(6.92)	12.6	(8.61)	16.0	(6.85)	12.54	(8.53)	15.44	(6.90)
2501-5000	16.9	(5.91)	18.4	(4.75)	17.3	(6.12)	18.8	(4.73)	16.52	(6.32)	17.69	(4.92)
>5000	26.5	(2.89)	22.4	(2.93)	27.3	(2.89)	22.8	(2.94)	24.20	(3.37)	20.71	(3.21)
Female												
$\leq$ 500	3.7	(7.02)	7.5	(5.19)	3.7	(7.04)	7.6	(5.04)	3.93	(7.28)	7.78	(5.09)
501-1000	4.4	(7.56)	8.4	(5.94)	4.4	(7.63)	8.4	(5.75)	4.57	(7.85)	8.57	(5.82)
1001-1500	5.3	(7.96)	9.5	(6.52)	5.3	(8.12)	9.5	(6.29)	5.46	(8.23)	9.54	(6.36)
1501-2500	6.7	(7.39)	11.0	(5.99)	6.8	(7.60)	10.9	(5.81)	6.78	(7.63)	10.75	(5.91)
2501-5000	9.2	(5.18)	12.9	(4.35)	9.2	(5.27)	12.8	(4.28)	8.93	(5.51)	12.32	(4.48)
>5000	14.4	(2.69)	15.8	(2.80)	14.6	(2.66)	15.5	(2.80)	13.01	(3.09)	14.43	(3.05)

Table 6: Asymptotically normal test statistic: B-W difference of VTTS

	Additive Logit	Reverse Logit	Probit
Male			
$\leq$ 500	-1.71*	-1.78*	-1.70*
501-100	-1.79*	-1.87*	-1.76*
1001-1500	-1.70*	-1.80*	-1.63
1501-2500	-1.16	-1.23	-1.08
2501-5000	0.30	-0.32	-0.26
>5000	0.34	-0.37	0.36
Female			
$\leq$ 500	-2.43**	-2.39**	-2.38**
501-1000	-2.61***	-2.56**	-2.53**
1001-1500	-2.60***	-2.53**	-2.49**
1501-2500	-2.06**	-2.00**	-1.96**
2501-5000	-1.09	-1.03	-1.06
>5000	-0.18	-0.12	-0.22
Key: ***,**,* = statistical significance at 1%, 5%, 10% level			

A third analysis relates to the evaluation of the predictive abilities of the models. This is also a test for similarities and differences among models estimated with B, W and B&W choices. We use the root likelihood as a scoring function to measure predictive accuracy. A scoring function is a measure of model prediction performance that relates the probabilistic model predictions to the stated choices (Louviere *et al.*, 2015). The root likelihood is preferred to the hit rate because it takes into account the probabilistic nature of the choice model, and it penalizes poor predictions in addition to rewarding accurate ones.

The root likelihood for prediction of B choices RL^B of the model estimated with choice data C is defined as

$$RL^B = \left\{ \prod_{i=1}^{N} \prod_{x \in X^i} \left[B^i_{X^i} \left(x^i, \mathbf{v}^i_C \right) \right]^{\Im^i_{BX^i}\left(x^i, \mathbf{v}^i_C\right)} \right\}^{1/N}, \tag{41}$$

where i is the observation superscript, N the number of observations in the sample, $\Im^i_{BX^i}\left(x^i, \mathbf{v}^i_C\right)$ the best choice indicator function for observation i and alternative x ($\Im^i_{BX^i}\left(x^i, \mathbf{v}^i_C\right) = 1$ if in observation i the alternative x is chosen as best,$= 0$ otherwise).

In the case where the model is estimated with B data, since

$$\ln\left(RL^B\right) = \frac{1}{N} \sum_{i=1}^{N} \sum_{x \in X^i} \Im^i_{BX^i}\left(x^i, \mathbf{v}^i_B\right) \ln B^i_{X^i}\left(x^i\right), \tag{42}$$

we also have

$$RL^B = e^{\frac{1}{N} LL^B}, \tag{43}$$

where LL^B is the total log-likelihood with best choices.

Similarly, the root likelihood for predicting W choices RL^W of the model estimated with choice data C is defined as

$$RL^W = \left\{ \prod_{i=1}^{N} \prod_{x \in X^i} \left[W^i_{X^i} \left(x^i, \mathbf{v}^i_C \right) \right]^{\Im^i_{WX^i}\left(x^i, \mathbf{v}^i_C\right)} \right\}^{1/N}, \tag{44}$$

where $\Im^i_{WX^i}\left(x^i, \mathbf{v}^i_C\right)$ is the worst choice indicator function for observation i and alternative x ($\Im^i_{WX^i}\left(x^i, \mathbf{v}^i_C\right) = 1$ if in observation i the alternative x is chosen as worst,= 0 otherwise), and LL^W is the total log-likelihood with worst choices.

In the case where the model is estimated with W data we have

$$RL^W = e^{\frac{1}{N} LL^W}, \tag{45}$$

where LL^W is the total log-likelihood with worst choices.

Table 7 shows, for the additive Logit, the reverse Logit and the Probit, the values of the root likelihood for predicting B and W choices with different types of choice data (B, W, B&W). The models estimated with B choice data best predict B choices (in the light of the highest root likelihood values), the models estimated with W choice data best predict W choices.

Notwithstanding the additional information, the models estimated with B&W data fail to improve prediction for either B and W choices relative to the models estimated with single-choice data.

In addition, the expectation that the models estimated with B&W data should provide better estimates, i.e. higher values of the t-statistic, than with B choice data or with W choice data only is not confirmed. These results are in agreement with Louviere *et al.* (2015, Ch. 14), who investigates Logit, using a reverse model for worst choices.

A fourth and final analysis relates to a comparison of the goodness of fit between the three models: additive Logit, reverse Logit and Probit. The analysis is carried out with reference to the AIC (Table 8). The model with the lowest AIC is the one with the best goodness of fit.

Table 7: Predictive accuracy of models estimated with different choice data: root likelihood

Estimation data	Additive Logit		Reverse Logit		Probit	
	Predict B choices RL^B	Predict W choices RL^W	Predict B choices RL^B	Predict W choices RL^W	Predict B choices RL^B	Predict W choices RL^W
B	.483°	.371	.484°	.402	.485°	.384
W	.454	.423°	.466	.421°	.416	.423°
B&W	.467	.419	.478	.418	.473	.416

° Highest value

Table 8: Goodness of fit of different models: AIC

Model	Predict B choices	Predict W choices	Predict B&W choices
additive Logit	855.5	1014.7°	1663.5
reverse Logit	857.5	1018.8	1621.6°
Probit	854.2°	1015.2	1635.5

° Lowest value

7. Conclusion

The paper has dealt with the estimation of RUMs using best, worst and best-worst data. Two distributions of the random terms are considered: Logit and Probit. Recent literature has provided closed-form expressions for best, worst and best-worst choice probabilities for Logit. The paper has proved that, in the case of Probit, worst choice probabilities are obtained from best choice probabilities by simply changing the sign of the systematic utilities. The paper has also proved the equality of best and worst choice probabilities between the additive and the reverse Probit. These properties of Probit are attributable to the symmetry of the univariate standard normal distribution.

A remarkable result of the paper is the proof, under mild conditions, of the strict log-concavity, with respect to the coefficients of the systematic utilities, of the likelihood functions associated with best, worst and best-worst choice probabilities in the case of Logit, and with best and worst choice probabilities in the case of Probit. The conditions boil down to the absence of perfect multicollinearity in the matrix of the alternative and individual characteristics, similar to linear regression models. The strict log-concavity of the likelihood associated with best-worst choice probabilities in Probit remains an open problem.

In the transportation mode choice example here, worst choice preferences, when compared with best choice preferences, show, in both Logit and Probit models, coefficient attenuation, and higher mean VTTS with larger confidence intervals. The results on VTTS suggest different best choice and worst choice preferences, even accounting for scale differences. This is, for Logit, in agreement with other authors (Ben-Akiva *et al.*, 1992; Louviere *et al.*, 2015; Giergiczny *et al.*, 2017. Only the result on the sign of the difference of the VTTS between best and worst choice data is conflicting with the paper by Giergiczny *et al.* (2017), who find lower mean VTTS with worst choice data. The finding on the lower statistical significance of differences in coefficients and VTTS in Probit than in Logit is in agreement with Nair *et al.* (2018b), who suggest rank-ordered Probit to be a superior option to rank-ordered Logit in terms of robustness of preferences across ranking depths.

The reasons for these results remain to be more extensively understood. Part of the explanation may be due to the behavioral shift: the respondent is usually more familiar with the best choice than with the worst choice.

In the empirical example, the benefit of collecting and using both best and worst choices as a data augmentation strategy for a best choice model is not confirmed. Therefore, at least in this case, the benefit of collecting best and worst choice data lies in the additional insights into the preferences of the individuals.

Data augmentation with worst choices is in all cases useful to estimate the probability of a given alternative ranking last. This would be of interest to marketing and the transportation field alike, because policy makers might wish to know if a given product or alternative had an unacceptable feature (when best-worst scaling is used for alternatives constituted by attributes of a single product or service). Future research might deal with an extension of this, estimating the probabilities, available in closed form from de Palma & Kilani (2015), of a given alternative ranking k-th, $k = 1, 2, \cdots, n$, in the choice set X (first, second,..., last).

Appendix A. Proofs of lemmas

Proof of Lemma 1. Using Theorem 10 in de Palma *et al.* (2017), for any $x, y \in X$, $x \neq y$, and any $\boldsymbol{v} \in \mathbb{R}^n$, the best and worst choice probabilities $BW_X(x, y; \boldsymbol{v})$ take the form given in (8) where

$$B_{X,Y}(x, y; \boldsymbol{v}) = \int_0^{\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 F_{\boldsymbol{\epsilon}}(u\mathbf{1}_X - t\mathbf{1}_Y - \boldsymbol{v})}{\partial t_x \partial t_y} du dt.$$

Using (3), we have

$$B_X(x; \boldsymbol{v} + t\mathbf{1}_Y) = \int_{-\infty}^{+\infty} \frac{\partial F_{\boldsymbol{\epsilon}}(u\mathbf{1}_X - \boldsymbol{v} - t\mathbf{1}_Y)}{\partial t_x} du.$$

Differentiation of the above equation with respect to v_y, yields

$$\frac{-\partial B_X(x; \boldsymbol{v} + t\mathbf{1}_Y)}{\partial v_y} = \int_{-\infty}^{+\infty} \frac{\partial^2 F_{\boldsymbol{\epsilon}}(u\mathbf{1}_X - \boldsymbol{v} - t\mathbf{1}_Y)}{\partial t_x \partial t_y} du.$$

Therefore, we get the alternative expression of $B_{X,Y}(x, y; \boldsymbol{v})$ provided in (9). □

Proof of Lemma 2. Differentiating (6), we obtain

$$W_X(x; \boldsymbol{v}) \nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \sum_{\{x\} \subseteq Y \subseteq X} (-1)^{|Y|-1} B_Y(x; \boldsymbol{v}) \nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x).$$

Dividing then both members of the above equation by $W_X(x; \boldsymbol{v})$, we obtain (12). Differentiating Expression (12), we get

$$\nabla_{\boldsymbol{\theta}}^2 \mathcal{L}_X^W(x) = \sum_{\{x\} \subseteq Y \subseteq X} \left\{ \omega_{X,Y}(x) \nabla_{\boldsymbol{\theta}}^2 \mathcal{L}_Y^B(x) + \left[\nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) \right]' \nabla_{\boldsymbol{\theta}} \omega_{X,Y}(x) \right\},$$

where $\nabla_{\boldsymbol{\theta}} \omega_{X,Y}(x)$ is the row gradient of $\omega_{X,Y}(x)$ with respect to $\boldsymbol{\theta}$. It satisfies

$$\nabla_{\boldsymbol{\theta}} \omega_{X,Y}(x) = \omega_{X,Y}(x) \nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) - \omega_{X,Y}(x) \nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x).$$

Combining the two above displayed equations and using (12), Eq. (13) is obtained. □

Appendix B. Proofs of propositions and theorems

Proof of Proposition 1. For the gradient computation, thanks to Lemma 2, Eq. (12), we have

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left(\boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_Y \right).$$

By expansion, and using the fact that $\sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) = 1$ and the Definition (29), we obtain (32).
For the expression of the Hessian matrix, we are using Lemma 2, Eq. (13). Notice first that using (26) and (28), for any $x \in Y \subseteq X$, we have

$$\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) + \left[\nabla_{\boldsymbol{\theta}} \mathcal{L}_Y^B(x) \right]^2 = H_Y(\boldsymbol{v}) + \left(\boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_Y \right)^2.$$

Therefore, using (13) and (32), we get

$$\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_X^W(x) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) H_Y(\boldsymbol{v}) + T_X(x; \boldsymbol{v}),$$

where

$$T_X(x; \boldsymbol{v}) \equiv \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left(\boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_Y \right)^2 - \left[\boldsymbol{\xi}_x - \overline{\boldsymbol{\xi}}_X^W(x) \right]^2.$$

The above expression can be rewritten as

$$T_X(x; \boldsymbol{v}) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left[\overline{\boldsymbol{\xi}}_Y - \overline{\boldsymbol{\xi}}_X^W(x) \right] \left[\overline{\boldsymbol{\xi}}_Y + \overline{\boldsymbol{\xi}}_X^W(x) - 2\boldsymbol{\xi}_x \right].$$

By expansion, we obtain

$$T_X(x; \boldsymbol{v}) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left[\overline{\boldsymbol{\xi}}_Y - \overline{\boldsymbol{\xi}}_X^W(x) \right] \overline{\boldsymbol{\xi}}_Y,$$

which can be rewritten as follows

$$T_X(x; \boldsymbol{v}) = \sum_{\{x\} \subseteq Y \subseteq X} \omega_{X,Y}(x) \left[\overline{\boldsymbol{\xi}}_Y - \overline{\boldsymbol{\xi}}_X^W(x) \right]^2,$$

obtaining the required expression. □

Proof of Proposition 2. The best choice probabilities, for any $\boldsymbol{v} \in \mathbb{R}^n$ and any $x \in X$, can be written as

$$B_X(x; \boldsymbol{v}) = \mathbb{P}\left(\bigcap_{z \in X \setminus \{x\}} \{\epsilon_z - \epsilon_x \leq v_x - v_z\} \right) = F_{\boldsymbol{\delta}}(\boldsymbol{w}), \tag{46}$$

where $F_{\boldsymbol{\delta}}(\cdot)$ is the c.d.f. of $\boldsymbol{\delta}$ whose components are $\epsilon_z - \epsilon_x,\ z \in X \setminus \{z\}$; the vector $\boldsymbol{w} \in \mathbb{R}^{n-1}$ has components: $w_z = v_x - v_z,\ z \in X \setminus \{x\}$.

Note that $\boldsymbol{w}$ is related to $\boldsymbol{d}$ by a reversible linear transformation. Indeed, setting $d_n = w_x = 0$, we have: $w_z = d_x - d_z$, $z \in X \setminus \{x\}$, and conversely: $d_z = w_n - w_z$, $z \in X \setminus \{n\}$. Strict concavity being preserved under any reversible linear transformation (Daganzo, 1979, Appendix D), strict log-concavity in $\boldsymbol{d}$ and $\boldsymbol{w}$ are equivalent. We therefore need to prove the strict long-concavity of $F_{\boldsymbol{\delta}}(\cdot)$.

In the case of the MNL, we have

$$F_{\boldsymbol{\delta}}(\boldsymbol{w}) = \left(\sum_{z \in X} e^{-w_z}\right)^{-1}, \ \boldsymbol{w} \in \mathbb{R}^{n-1},$$

which is the distribution of the classical $(n-1)$-variate logistic distribution (see Arnold, 1992, p. 238). Since,

$$\ln F_{\boldsymbol{\delta}}(\boldsymbol{w}) = -\ln\left(\sum_{z \in X} e^{-w_z}\right), \ \boldsymbol{w} \in \mathbb{R}^{n-1},$$

we get, for $z_1, z_2 \in X \setminus \{x\}$, the following second-order partial derivatives

$$\frac{\partial^2 \ln F_{\boldsymbol{\delta}}(\boldsymbol{w})}{\partial w_{z_1} \partial w_{z_2}} = \frac{e^{-w_{z_1}}}{\left(\sum_{z \in X} e^{-w_z}\right)^2} \begin{cases} -\sum_{z \in X \setminus \{z_1\}} e^{-w_z}, & z_2 = z_1; \\ e^{-w_{z_2}}, & z_2 \neq z_1. \end{cases}$$

The Hessian of $\ln F_{\boldsymbol{\delta}}(\cdot)$ is a symmetric matrix which has a negative dominant diagonal since

$$\sum_{z \in X \setminus \{z_1\}} e^{-w_z} > \sum_{z \in X \setminus \{z_1, n\}} e^{-w_z}, \ z_1 \in X \setminus \{x\}.$$

Accordingly (see, e.g., Mas-Colell, 1985), the Hessian of $\ln F_{\boldsymbol{\delta}}(\cdot)$ is negative definite so that $F_{\boldsymbol{\delta}}(\cdot)$ is strictly log-concave.

Similarly, for the worst choice, we have

$$W_X(x; \boldsymbol{v}) = \mathbb{P}\left(\bigcap_{z \in X \setminus \{x\}} \{\epsilon_x - \epsilon_z \leq v_z - v_x\}\right) = F_{-\boldsymbol{\delta}}(-\boldsymbol{w}), \tag{47}$$

where $F_{-\boldsymbol{\delta}}(\cdot)$ is the c.d.f. of $-\boldsymbol{\delta}$. We need to prove that $F_{-\boldsymbol{\delta}}(\cdot)$ is strictly log-concave. As well known, the p.d.f. $f_{-\boldsymbol{\delta}}(\cdot)$ of $-\boldsymbol{\delta}$ and the p.d.f. $f_{\boldsymbol{\delta}}(\cdot)$ of $\boldsymbol{\delta}$ are related by $f_{-\boldsymbol{\delta}}(-\boldsymbol{w}) = f_{\boldsymbol{\delta}}(\boldsymbol{w})$. The p.d.f. of the classical $(n-1)$-variate logistic distribution satisfies

$$f_{\boldsymbol{\delta}}(\boldsymbol{w}) = (n-1)! e^{-\sum_{z \in X} w_z} \left(\sum_{z \in X} e^{-w_z}\right)^{-n},$$

so that

$$\ln f_{\boldsymbol{\delta}}(\boldsymbol{w}) = \ln\left[(n-1)!\right] - \sum_{z \in X} w_z + n \ln F_{\boldsymbol{\delta}}(\boldsymbol{w}).$$

The RHS is the sum of an affine function and a strictly concave expression in $\boldsymbol{w}$ (recall that $F_{\boldsymbol{\delta}}(\cdot)$ is stricly log-concave), so that $f_{\boldsymbol{\delta}}(\cdot)$ (and therefore $f_{-\boldsymbol{\delta}}(\cdot)$) is strictly log-concave. Applying (Prékopa, 1973, Theorem 5), $F_{-\boldsymbol{\delta}}(\cdot)$ is also strictly log-concave, so that $W_X(x;\boldsymbol{v})$ is strictly log-concave in $\boldsymbol{d}$. □

Proof of Theorem 1. Consider one observation (the index i is omitted for simplicity). Based on Proposition 2, the best choice probabilities $B(x;\boldsymbol{v})$ and the worst choice probabilities $W(x;\boldsymbol{v})$ are strictly log-concave in $\mathbf{d}$ for any $x \in X$. We have $d_z = w_n - w_z$, $z \in X \setminus \{n\}$, with $w_z = v_x - v_z$. Therefore, $d_z = v_z - v_n$, $z \in X \setminus \{n\}$. Recall that the systematic utilities are assumed linear in the coefficients: $v_z = \boldsymbol{\xi}_z\boldsymbol{\theta}$, $z \in X$.

Based on Daganzo (1979, Appendix D), if $F(\mathbf{x})$ is strictly concave in $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{A}$ is an arbitrary $m \times n$ matrix, and $\mathbf{y}$ is a m-dimensional row vector, then $G(\mathbf{y}) = F(\mathbf{yA})$ is strictly concave in $\mathbf{y}$. Therefore, the natural logarithm of $B(x;\boldsymbol{v})$ and of $W(x;\boldsymbol{v})$ is strictly concave in $\boldsymbol{\theta}$. The strict log-concavity of the total likelihood follows from preservation of strict log-concavity under product.

If the matrix $\mathcal{M}$ has not full rank, i.e. if at least one column is linear combination of the others, it is not possible to estimate the independent impact of the change of all the alternative and individual characteristics. This circumstance must be excluded, or, as underlined by Wedderburn (1976), there would be non-uniqueness. An example clarifies this point. Assume we have for all observations (related index omitted) $\xi_1 = \alpha\xi_2$. Clearly $\mathcal{M}$ has not full rank. Then we have identical values of probabilities for different combinations of θ_1 and θ_2, i.e. non-uniqueness. This is because we have:

$$v_z = \alpha\xi_{2,z}\theta_1 + \xi_{2,z}\theta_2 + \sum_{h\neq 1,2} \xi_{h,z}\theta_h,\ z \in X$$

Computation of utility, and probability, depends on $\alpha\theta_1 + \theta_2$. There are infinite combinations of θ_1 and θ_2 yielding the same value of $\alpha\theta_1 + \theta_2$. □

Proof of Proposition 3. In the light of Eqs (2) and (5), we only need to prove that the vector $-\boldsymbol{\epsilon}$, which is MVN since it is an affine transformation of $\boldsymbol{\epsilon}$, has zero mean and covariance matrix $\boldsymbol{\Sigma}$.
Recall the following property (Brockwell & Davis, 1991, Proposition 1.6.1.): if $\mathbf{a}$ is a m-component column vector, $\mathbf{B}$ is a $m \times n$ matrix, and $\mathbf{X}$ a n-component random vector with mean $\mathbb{E}(\mathbf{X})$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{XX}}$, then the random vector $\mathbf{Y} = \mathbf{a} + \mathbf{BX}$, has mean $\mathbb{E}(\mathbf{Y}) = \mathbf{a} + \mathbf{B}\mathbb{E}(\mathbf{X})$, and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{YY}} = \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{XX}}\mathbf{B}'$. We only need to apply this property to $\boldsymbol{\epsilon}$, with $\mathbf{a} = \mathbf{0}$ and $\mathbf{B} = -\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. □

Proof of Proposition 4. The proof goes along the same lines as the proof of Proposition 2. We use Eq. (46), which holds for any ARUM. In the Probit case, the random vector $\boldsymbol{\delta}$ is a MVN with p.d.f.

$$f_{\boldsymbol{\delta}}(\boldsymbol{w}) = \frac{1}{(2\pi)^{(n-1)/2}\left[\det(\boldsymbol{\Delta})\right]^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{w}'\boldsymbol{\Delta}^{-1}\boldsymbol{w}\right),$$

where $\boldsymbol{w} \in \mathbb{R}^{n-1}$ and $\boldsymbol{\Delta}$ is the covariance matrix of $\boldsymbol{\delta}$, provided that this matrix is non-singular, which is proved below. Therefore, we get

$$\ln f_{\boldsymbol{\delta}}(\boldsymbol{w}) = -\frac{1}{2}\boldsymbol{w}'\boldsymbol{\Delta}^{-1}\boldsymbol{w} - \frac{n-1}{2}\ln(2\pi) - \frac{1}{2}\ln[\det(\boldsymbol{\Delta})].$$

The first term of the RHS is a symmetric bilinear form with matrix $(-1/2)\boldsymbol{\Delta}^{-1}$, the remaining terms being constants. Since the covariance matrix $\boldsymbol{\Delta}$ is positive definite, $\boldsymbol{\Delta}^{-1}$ is also positive definite (Cambini & Martein, 2009, page 233). Hence $(-1/2)\boldsymbol{\Delta}^{-1}$ is negative definite so that $f_{\boldsymbol{\delta}}(\cdot)$ is stritly log-concave. By Theorem 5 of Prékopa (1973), $F_{\boldsymbol{\delta}}(\cdot)$ also is strictly log-concave implying that $B_X(x, \mathbf{v})$ it is also strictly log-concave in $\mathbf{d}$.

For the worst choice probability in the Probit case, using (47), recall that $W_X(x;\boldsymbol{v}) = F_{-\boldsymbol{\delta}}(-\boldsymbol{w})$. Since $F_{-\boldsymbol{\delta}}(\cdot) = F_{\boldsymbol{\delta}}(\cdot)$, the worst choice probabilities also are strictly long-concave.

We now prove that $\boldsymbol{\Delta}$ is non-singular, therefore positive definite, if $\boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{\epsilon}$, is positive definite. W.l.o.g., and for the sake of notational simplicity, we set $x = n$. We denote by $\Delta_{z_1z_2}$ (resp. $\sigma_{z_1z_2}$) the (z_1, z_2)-th element of $\boldsymbol{\Delta}$ (resp. $\boldsymbol{\Sigma}$), for z_1, z_2 belonging to $X \setminus \{n\}$ (resp. X). They are related by

$$\Delta_{z_1z_2} = \sigma_{z_1z_2} - \sigma_{z_1n} - \sigma_{z_2n} + \sigma_{nn},\ z_1, z_2 = 1, \ldots, n-1.$$

For any $\mathbf{w}$ non-null vector of $\mathbb{R}^{n-1}$, let define

$$\mathbf{w}'\boldsymbol{\Delta}\mathbf{w} \equiv \sum_{z_1=1}^{n-1}\sum_{z_2=1}^{n-1} \Delta_{z_1z_2} w_{z_1} w_{z_2}.$$

Let $w_n = -\sum_{z=1}^{n-1} w_z$. Combining the two above expressions, we get

$$\begin{aligned} \mathbf{w}'\boldsymbol{\Delta}\mathbf{w} &= \sum_{z_1=1}^{n-1}\sum_{z_2=1}^{n-1} \sigma_{z_1z_2} w_{z_1} w_{z_2} \\ &+ 2\sum_{z=1}^{n-1} \sigma_{zn} w_z w_n + \sigma_{nn} w_n^2. \end{aligned}$$

Therefore,

$$\mathbf{w}'\boldsymbol{\Delta}\mathbf{w} = \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x},$$

where $\mathbf{x} = (\mathbf{w}, w_n) \in \mathbb{R}^n$. Since $\mathbf{x}$ is non-null and $\boldsymbol{\Sigma}$ is definite positive, we have $\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x} > 0$ so that $\mathbf{w}'\boldsymbol{\Delta}\mathbf{w} > 0$. Hence $\boldsymbol{\Delta}$ is definite positive and therefore non-singular. □

Proof of Theorem 2. Consider one observation (the index i is omitted for simplicity). Based on Proposition 4, the best choice probabilities $B(x;\boldsymbol{v})$ and the worst choice probabilities $W(x;\boldsymbol{v})$ are strictly log-concave in $\mathbf{d}$ for any $x \in X$. We have $d_z = w_n - w_z$, $z \in X \setminus \{n\}$, with $w_z = v_x - v_z$. Therefore,

$d_z = v_z - v_n,\ z \in X \setminus \{n\}$. Recall that the systematic utilities are assumed linear in the coefficients: $v_z = \boldsymbol{\xi}_z\boldsymbol{\theta},\ z \in X$.

Based on (Daganzo, 1979, Appendix D), if $F(\mathbf{x})$ is strictly concave in $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{A}$ is an arbitrary $m \times n$ matrix, and $\mathbf{y}$ is a m-dimensional row vector, then $G(\mathbf{y}) = F(\mathbf{yA})$ is strictly concave in $\mathbf{y}$. Therefore, the natural logarithm of $B(x; \boldsymbol{v})$ and of $W(x; \boldsymbol{v})$ is strictly concave in $\boldsymbol{\theta}$. The strict log-concavity of the total likelihood follows from preservation of strict log-concavity under product.

If the matrix $\mathcal{M}$ has not full rank, i.e. if at least one column is linear combination of the others, it is not possible to estimate the independent impact of the change of all the alternative and individual characteristics. This circumstance must be excluded, or, as underlined by (Wedderburn, 1976), there would be non-uniqueness. An example clarifies this point. Assume we have for all observations (related index omitted) $\xi_1 = \alpha\xi_2$. Clearly $\mathcal{M}$ has not full rank. Then we have identical values of probabilities for different combinations of θ_1 and θ_2, i.e. non-uniqueness. This is because we have:

$$v_z = \alpha\xi_{2,z}\theta_1 + \xi_{2,z}\theta_2 + \sum_{h \neq 1,2} \xi_{h,z}\theta_h,\ z \in X$$

Computation of utility, and probability, depends on $\alpha\theta_1 + \theta_2$. There are infinite combinations of θ_1 and θ_2 yielding the same value of $\alpha\theta_1 + \theta_2$. □

Proof of Proposition 5. The proof is along the same lines of the proof of Proposition 3. It is left to the reader. □

Acknowledgments

The authors would like to thank the Associate Editor for helpful comments on an earlier version of the paper. The authors are also grateful to Alfredo Donno, Julien Monardo and to the participants of the 2018 Annual Conference of the International Transportation Economics Association, Hong Kong.

References

Amemiya, T. 1985. *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

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

Arnold, B.C. 1992. Multivariate logistic distributions. *Chap. 11, pages 237–261 of:* Balakrishnan, N. (ed), *Handbook of the Logistic Distribution.* Statistics: Textbooks and Monographs. New York: Marcel Dekker.

Beck, M.J., Rose, J.M., & Greaves, S.P. 2017. I can't believe your attitude: a joint estimation of best worst attitudes and electric vehicle choice. *Transportation*, **44**(4), 753–772.

Beggs, S., Cardell, S., & Hausman, J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics*, **17**(1), 1–19.

Ben-Akiva, M., & Lerman, S.R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge MA: The MIT Press.

Ben-Akiva, M., Morikawa, T., & Shiroishi, F. 1992. Analysis of the reliability of preference ranking data. *Journal of Business Research*, **24**(2), 149–164.

Brockwell, P.J., & Davis, R.A. 1991. *Time Series: Theory and Methods*. New York: Springer.

Cambini, A., & Martein, L. 2009. *Generalized Convexity and Optimization: Theory and Applications*. Berlin: Springer-Verlag.

Daganzo, C. 1979. *Multinomial Probit: the Theory and its Application to Demand Forecasting*. New York: Academic Press.

de Palma, A., & Kilani, K. 2015. *Ordered choice probabilities in random utility models*. Ecole Polytechnique. Cahier n° 2015-04.

de Palma, A., Kilani, K., & Laffond, G. 2017. Relations between best, worst, and best-worst choices for random utility models. *Journal of Mathematical Psychology*, **76**(Part A), 51–58.

Finn, A., & Louviere, J.J. 1992. Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing*, **11**(2), 12–25.

Flynn, T.N., & Marley, A.A. 2014. Best-worst scaling: theory and methods. *Pages 178–199 of:* Hess, S., & Daly, A. (eds), *Handbook of Choice Modelling*. Cheltenham UK: Edward Elgar.

Fok, D., Paap, R., & Van Dijk, B. 2012. A rank ordered logit model with unobserved heterogeneity in ranking capabilities. *Journal of Applied Econometrics*, **27**(5), 831–846.

Gatta, V., Marcucci, E., & Scaccia, L. 2015. On finite sample performance of confidence intervals methods for willingness to pay measures. *Transportation Research Part A: Policy and Practice*, **82**, 169–192.

Giergiczny, M., Dekker, T., Hess, S., & Chintakayala, P. 2017. Testing and stability of utility parameters in repeated best, repeated best-worst, and one-off best-worst studies. *European Journal of Transport and Infrastructure Research*, **17**(4), 457–476.

Greene, W.H. 2012. *LIMDEP Version 10 Reference Guide*. Econometric Software, Inc.

Hajivassiliou, V.A., & Ruud, P.A. 1994. Classical estimation methods for LDV models using simulation. *Pages 2383–2441 of:* Engle, R.F., & McFadden, D. (eds), *Handbook of Econometrics*, vol. 4. Elsevier.

Louviere, J.J., & Islam, T. 2008. A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best-worst scaling. *Journal of Business Research*, **61**(9), 903–911.

Louviere, J.J., & Woodworth, G. 1990. *Best-worst scaling: a model for largest difference judgments.* Working Paper. University of Alberta.

Louviere, J.J., Flynn, T.N., & Marley, A.A. 2015. *Best-Worst Scaling: Theory, Methods and Applications.* Cambridge UK: Cambridge University Press.

Luce, R.D., & Suppes, P. 1965. Preference, utility and subjective probability. *Chap. 19, pages 251–410 of:* Luce, R.D., Bush, R.R., & Galanter, E. (eds), *Handbook of Mathematical Psychology*, vol. III. New York: Wiley and Sons.

Marley, A.A., & Flynn, T.N. 2015. Best and worst scaling: theory and application. *Pages 548–552 of:* Wright, J.D. (ed), *International Encyclopedia of the Social and Behavioural Sciences, 2nd edition*, vol. 2. Elsevier Science.

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

Mas-Colell, A. 1985. *The Theory of General Equilibrium: a Differentiable Approach.* Cambridge UK: Cambridge University Press.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. *Pages 105–142 of:* Zarembka, P. (ed), *Frontiers in Econometrics.* New York NY: Academic Press.

McFadden, D. 1981. Econometric Models of Probabilistic Choice. *Pages 198–272 of:* Manski, C.F., & McFadden, D. (eds), *Structural analysis of discrete dara with econometric applications.* Cambridge (Mass.): MIT Press.

Nair, G.S., Astroza, S., Bhat, C.R., Khoeini, S., & Pendyala, R.M. 2018a. An application of a rank-ordered probit modeling approach to understanding level of interest in autonomous vehicles. *Transportation.*

Nair, G.S., Bhat, C.R., Pendyala, R.M., Loo, B.P.Y., & Lam, W.H.K. 2018b. *On the use of probit based models for ranking data analysis.*

Prékopa, A. 1973. On logarithmic concave measures and functions. *Acta Scientiarum and Mathematicarum*, **34**(1-4), 335–343.

Scarpa, R., Notaro, S., Louviere, J.J., & Raffaelli, R. 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in Alpine grazing commons. *American Journal of Agricultural Economics*, **93**(3), 813–828.

S.P., Anderson, de Palma A., & J.-F., Thisse. 1992. *Discrete choice theory of product differentiation*. Cambridge MA: The MIT Press.

Swait, J., & Louviere, J.J. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, **30**(3), 305–314.

Vann Ophem, H., Stam, P., & van Praag, B. 1999. Multichoice logit: modeling incomplete preference rankings of classical concerts. *Journal of Business & Economic Statistics*, **17**(1), 117–128.

Wedderburn, R.W.M. 1976. On the existence and uniqueness of the maximum likelihood estimates of certain Generalised Linear Models. *Biometrika*, **63**, 27–32.

Yan, J., & Yoo, H. 2014. *The seeming unreliability of rank-ordered data as a consequence of model specification*. MPRA Paper No 56285.