Transportation Research Part B 121 (2019) 74–91

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A new family of qualitative choice models: An application of reference models to travel mode choice

Hélène Bouscasse[a], Iragaël Joly[b], Jean Peyhardi[c,*]

[a] CESAER, Agrosup Dijon, INRA, Université Bourgogne Franche-Comté, Dijon, France
[b] Université Grenoble Alpes, CNRS, INRA, Grenoble INP, GAEL, Grenoble 38000 France
[c] IGF, Univ. Montpellier, CNRS, INSERM, Montpellier, France

ARTICLE INFO

Article history:
Received 5 April 2018
Revised 21 December 2018
Accepted 28 December 2018
Available online 16 January 2019

MSC:
C25
D04
R41

Keywords:
Discrete choice model
Generalized linear model
Link function
Individual choice behavior

ABSTRACT

This paper considers the recently introduced family of reference models dedicated to non-ordered alternatives. The link function of reference models is that of the multinomial logit model (MNL) replacing the logistic cumulative distribution function (cdf) by other cdfs (e.g., Gumbel, Student). We determine all usual economic outputs (willingness-to-pay, elasticities,...). We also show that the IIA property generally does not hold for this family of models, because of their noninvariance to the alternative chosen as a reference. We estimate and compare five reference models to the MNL on a travel mode-choice survey: according to the chosen cdf, reference models lead to a better fit and retrieve consistent economic outputs estimations even when there is a high unobserved heterogeneity.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Discrete choice models (DCMs) are a valuable tool for both understanding how choices are made and deriving behavioral outputs for economic analysis and valuation. DCMs have been applied in a variety of fields (transportation, energy, food consumption, marketing, and so on) with varied objectives (preferences analysis, value elicitation, estimation of demand elasticities, market share prediction, and so on). They are mainly applied to model consumer choice behavior but also applied to industrial economics. For example, Billot and Thisse (1995) use DCM to analyze product differentiation and agents selection within an organization.

Some DCMs are derived from the random utility maximization (RUM) principle such as the family of generalized extreme value (GEV) models introduced by McFadden (1978), which includes the MNL. However, the family of DCMs is principally composed of generalized linear models (GLMs) for discrete response variable, also including the MNL. A non-exhaustive overview of DCMs is presented in Fig. 1. Discrete response variables can be differentiated into count and qualitative responses. On the one hand, Poisson regression models are classically used for count data (Cameron and Trivedi, 1986). On the other hand, different models are used for qualitative choices according to the ordering assumption among the alternatives. Regression models appropriate for ordered alternatives include the cumulative (McCullagh, 1980), sequential (Tutz, 1991) and

* Corresponding author.
E-mail addresses: helene.bouscasse@inra.fr (H. Bouscasse), iragael.joly@grenoble-inp.fr (I. Joly), jean.peyhardi@umontpellier.fr (J. Peyhardi).

https://doi.org/10.1016/j.trb.2018.12.010
0191-2615/© 2019 Elsevier Ltd. All rights reserved.

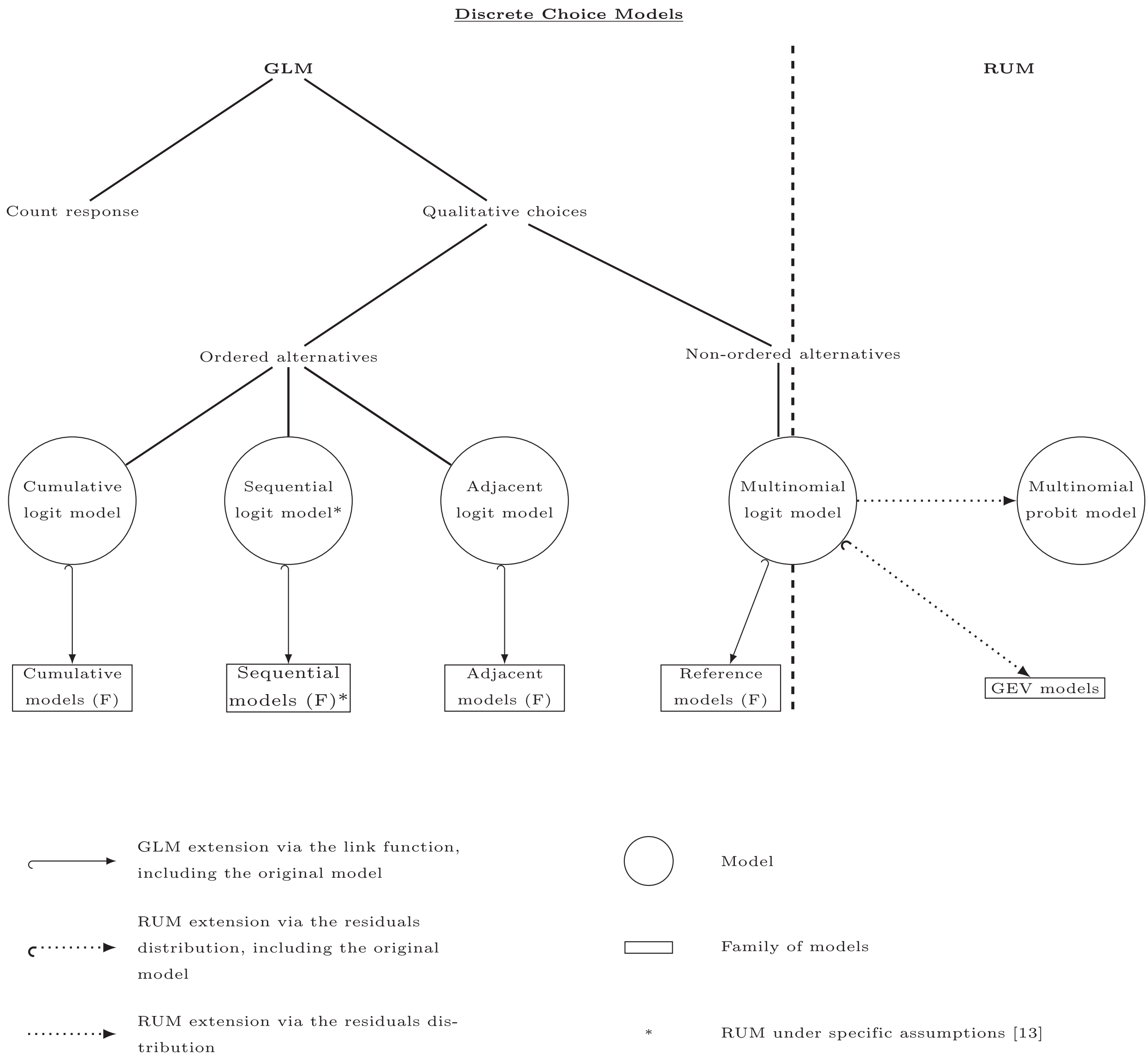


Fig. 1. Reference models position among discrete choice models.

adjacent (Goodman, 1983) logit model. They have been used, for instance, to explore the dynamic of credit rating matrices (Feng et al., 2008), to explain employment stability (Kahn and Morimune, 1979) or for partial credit scoring (Masters, 1982). For these three kinds of ordered logit models some extensions have been introduced using a cumulative distribution function (cdf) other than the logistic one (Amemiya, 1981; Fullerton, 2009). The best known of them is the ordered probit model, which is an extension of the cumulative logit model, defined with the normal cdf (McKelvey and Zavoina, 1975).

In the case of non-ordered alternatives, such extensions have only recently been proposed for the MNL leading to the family of reference models (Peyhardi et al.). It should be noted that the multinomial probit model is not such an extension, despite its name. It is defined as a RUM model with multivariate Gaussian residual distribution whereas the cumulative, sequential[1] and adjacent models do not strictly satisfy the RUM principle (Small, 1987) and are viewed as GLMs. Otherwise, reference models take advantage of the GLM framework. The simple Newton-Raphson or Fisher scoring algorithm can be used, implying a fast estimation procedure compared to the complex estimation of the multinomial probit model; see Peyhardi et al. for details. In the GLM framework, the link function relates the probabilities of alternatives to the linear predictors (depending on individual and alternative-specific variables). The link function of reference models is the multinomial logit link function, except that the logistic cdf is replaced by other cdfs F, such as the normal, the Gompertz or the

[1] The sequential logit model can be viewed as a RUM model under specific assumptions (Sheffi, 1979).

Student cdfs. Adjusting the tail-weight and asymmetry of distributions and relaxing assumptions of the MNL (for example, the IIA) model may markedly improve the model fit (Peyhardi et al.) and may have behavioral and economic implications. The MNL has been widely used to perform economic analysis (Small and Rosen, 1981), design public policies or adapt marketing strategies (Baltas and Doyle, 2001) in a variety of fields (transportation, energy, food consumption, and so on). The use of reference models for discrete choice modeling may therefore have strong implications if they provide results that are contrasted with the MNL in terms of data fit, preferences analysis, value elicitation or demand elasticities. The aim of this article is to derive economic outputs (preferences structure, marginal effects, willingness-to-pay, and elasticities) from the reference models and to empirically compare these new qualitative choice models with the MNL. To evaluate the added value that reference models offer, in terms of statistical, behavioral and economic results, the empirical application uses a recent stated preferences (SP) survey that explores the determinants of travel mode choice. Mode choice is particularly appropriate for such an application since it is a well documented decision process (see De Witte et al., 2013, for a literature review) which has already been the support for the development of DCMs.

The paper is organized as follows. Section 2 presents the family of reference models in the context of DCM and derives economic outputs. Section 3 presents the data, the design and specification of the models and discusses the results. Section 4 concludes.

2. Generalized linear models for non-ordered choices

2.1. Logit models

Let $\pi_j^{\{n\}}$ denote the probability that individual n chooses the alternative j among the alternatives $1, \ldots, J$. Let $x^{\{n\}}$ and $\omega^{\{n\}} = \left(\omega_j^{\{n\}}\right)_{j=1,\ldots,J}$ respectively denote the vector of individual and alternative-specific variables associated with the individual n. For the sake of simplicity, the individual index n will be suppressed in the following without loss of generality. The logit model is summarized by the $J-1$ equations

$$\pi_j = \frac{\exp(\eta_j)}{1 + \sum_{k=1}^{J-1} \exp(\eta_k)}, \tag{1}$$

for $j = 1, \ldots, J-1$. Depending on the form of the linear predictors η_j, we obtain different logit models:

- $\eta_j^{(1)} = \alpha_j + x^T \delta_j$. Individual variables x are used with $J-1$ different slopes δ_j.
- $\eta_j^{(2)} = \alpha_j + \tilde{\omega}_j^T \gamma$ where $\tilde{\omega}_j = \omega_j - \omega_J$. Alternative-specific variables ω_j are used with common slope γ. This is the conditional logit model introduced by McFadden (1973).
- $\eta_j^{(3)} = \alpha_j + x^T \delta_j + \tilde{\omega}_j^T \gamma$. This is a combination of the two previous parameterizations leading to the widely used MNL.

2.2. Reference models

All the classical regression models for categorical responses described by Tutz (2012) and Agresti (2013) share the generic equations (Peyhardi et al.):

$$r_j(\pi) = F(\eta_j) \tag{2}$$

for $j = 1, \ldots, J-1$, where r is a diffeomorphism[2] from the simplex (corner of hypercube) to an open subset of the hypercube, π is the vector of probabilities $(\pi_1, \ldots, \pi_{J-1})^T$ and F is a continuous and strictly increasing cdf. In this framework, the Eq. (1) of the logit models are equivalent to

$$\frac{\pi_j}{\pi_j + \pi_J} = \frac{\exp(\eta_j)}{1 + \exp(\eta_j)},$$

for $j = 1, \ldots, J-1$. We recognize the logistic cdf (on the right) and the recently introduced reference ratio (on the left). The alternative J is considered as the reference alternative and thus $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$. Moreover, the vector of predictors $\eta = (\eta_1, \ldots, \eta_{J-1})^T$ can be expressed as the product of the design matrix Z and the vector of parameters. The three previous parameterizations of linear predictors can thus be summarized by the three following design matrices:

$$Z^{(1)} = \begin{pmatrix} 1 & & & x^T & & \\ & \ddots & & & \ddots & \\ & & 1 & & & x^T \end{pmatrix},$$

[2] Recall that a differentiable map is a diffeomorphism if it is a bijection and its inverse is differentiable. The inverse differentiability is required for Fisher's scoring algorithm.

Table 1
Characteristics of the cumulative distribution functions $F \in \mathfrak{F}^0$.

Name	F(x)	Shape parameter	Symmetric	(m, s) such that $F_{m,s} \in \mathfrak{F}_{q_{0.95}}$	
logistic	$\frac{\exp(x)}{1+\exp(x)}$	No	Yes	$m = 0$	$s = 1$
normal	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$	No	Yes	$m = 0$	$s = 1.79$
Laplace	$\begin{cases} \frac{1}{2}\exp(x) & \text{if } x < 0 \\ 1-\frac{1}{2}\exp(-x) & \text{if } x \geq 0 \end{cases}$	No	Yes	$m = 0$	$s = 1.279$
Cauchy	$\frac{1}{2} + \frac{1}{\pi}\arctan(x)$	No	Yes	$m = 0$	$s = 0.466$
Gumbel	$\exp\{-\exp(-x)\}$	No	No	$m = -0.414$	$s = 1.131$
Gompertz	$1 - \exp\{-\exp(x)\}$	No	No	$m = 0.737$	$s = 2.012$
Student	$\frac{1}{2} + x\Gamma\left(\frac{1+\nu}{2}\right) \frac{{}_2F_1\left(\frac{1}{2}, \frac{1+\nu}{2}, \frac{3}{2}, -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$	ν (df)	Yes	$m = 0$	$s = 0.0123$ ($\nu = 0.35$)

$$Z^{(2)} = \begin{pmatrix} 1 & & & \tilde{\omega}_1^T \\ & \ddots & & \vdots \\ & & 1 & \tilde{\omega}_{J-1}^T \end{pmatrix},$$

$$Z^{(3)} = \begin{pmatrix} 1 & & & x^T & & & \tilde{\omega}_1^T \\ & \ddots & & & \ddots & & \vdots \\ & & 1 & & & x^T & \tilde{\omega}_{J-1}^T \end{pmatrix}.$$

Finally, the three logit models are fully specified by the (reference, logistic, $Z^{(i)}$) triplets with $i = 1, 2, 3$.

Other models can be derived from Eq. (2) by choosing alternative ratios $(r_j)_{j \neq J}$ and/or alternative cdfs F. So far, four ratios have been highlighted: the reference, adjacent, cumulative and sequential ratios. The last three are appropriate for ordered alternatives, since the models generated are generally not invariant under a permutation of alternatives. On the contrary, the reference models are invariant under all permutations that fix the reference alternative and are thus appropriate for non-ordered alternatives; see Peyhardi et al. for details about invariance properties. Although the reference ratio is mandatory for non-ordered choices, the logistic cdf is not. The cdf F only needs to be strictly increasing for the interpretation of parameters and differentiable for their estimation. The set of such cdfs is denoted by $\mathfrak{F}$. In application, this set will be restricted to the logistic, normal, Laplace, Cauchy, Gumbel, Gompertz and Student cdfs (i.e., all the most usual distributions of $\mathfrak{F}$) and will be denoted by $\mathfrak{F}^0$. The characteristics of these distributions are summarized in Table 1. Moreover, $\mathfrak{F}^*$ (respectively $\mathfrak{F}^{0,*}$) will denote the same set, excluding the logistic cdf. Cdfs differ in their symmetry (symmetric or asymmetric distributions) and in their tails (heavy or thin tails), each allowing a different fit of observed probabilities according to the predictor values. For example, heavy-tailed distributions, such as the Cauchy or Student cdfs, permit an abrupt transition of the response behavior around a threshold value of the predictor. The asymmetry of the Gumbel and Gompertz distributions allows an asymmetry in the response behavior around the switching point. The selection of the appropriated cdf F may markedly improve the model fit and the accuracy of the estimated economic outputs. Finally, we propose to use the family of reference models specified by the triplets (reference, F, $Z^{(i)})_{F \in \mathfrak{F}^0}$ with $i = 1, 2, 3$, containing the usual logit models when F is the logistic cdf.

2.3. Normalization of parameter estimates via the cdf F

As noted by Tutz (1991), “distribution functions generate the same model if they are connected by a linear transformation”. If the connection is made through a location parameter m and a scale parameter s such that $F_{m,s}(z) = F\{(z - m)/s\}$, we have

$$F_{m,s}(\eta'_j) = F\left(\frac{\eta'_j - m}{s}\right) = F\left\{\frac{\alpha'_j - m}{s} + x^T \frac{\delta'_j}{s} + (\omega_j - \omega_J)^T \frac{\gamma'}{s}\right\},$$

for $j = 1, \ldots, J-1$. An equivalent model is obtained using the reparametrization $\alpha'_j = s\alpha_j + m$, $\delta'_j = s\delta_j$ and $\gamma' = s\gamma$. Therefore, for each distribution, only one representative element F_{m_0,s_0} has to be chosen among the class of cdfs $(F_{m,\,s})_{m,\,s>0}$. In the context of binary regression, Tutz (2012) proposed the transformation of F so that the mean and the variance according to different distributions are the same, using the logistic cdf as the reference (mean and variance of the logistic cdf are 0 and $\pi^2/3$). However, this method cannot be used with the Cauchy cdf since the mean and variance do not exist. More generally, the mean and variance of the Student distribution do not exist when $\nu < 1$ and $\nu < 2$ respectively. Therefore, we propose to normalize the different reference models via the location and scale parameters of the cdf F. To do this, two real points a

and b are chosen such that all cdfs in $\mathfrak{F}$ have the same values $F(a)$ and $F(b)$. The logistic cdf is naturally proposed as the reference cdf since it corresponds to the canonical GLM; see Appendix A for details about the normalization of parameters.

2.4. The case of the Student distribution

We have seen that a change in location and/or scale parameters does not impact the loglikelihood of a reference model. Therefore, only one representative element F_{m_0,s_0} is chosen for each distribution: the logistic, normal, Laplace, Cauchy, Gumbel and Gompertz distributions. It is quite different for the Student distribution because of its additional parameter: the degree of freedom (df). Indeed, Student distributions with different degrees of freedom are not connected by a linear transformation and thus lead to different loglikelihood maxima. Since the df has to be estimated, we are interested in the concavity of the loglikelihood according to the parameter $\nu \in (0, \infty)$. This concavity is not theoretically demonstrated but is empirically verified; see Appendix B for more details. It should be recalled that the Cauchy and Student with $\nu = 1$ are the same distribution. Therefore, the (reference, Student$_{\hat{\nu}}$, Z) model will always obtain a higher loglikelihood than the (reference, Cauchy, Z) model. However, this latter model will be used because it is more parsimonious.

2.5. Independence of irrelevant alternatives

Unlike the logit models, the (reference, F, $Z)_{F\in\mathfrak{F}^*}$ are not invariant under the transposition of the reference alternative. Thus, the choice of the reference alternative will impact the parameter estimation and also the different economic outputs, such as the elasticities for instance. This dependence on the reference alternative allows us to partially relax the IIA property. The IIA property holds for the $J-1$ ratios of probabilities π_j/π_J since

$$\frac{\pi_j}{\pi_J} = \frac{F}{1-F}(\eta_j)$$

for all $j \neq J$. However, for two non-reference alternatives $j \neq J$ and $k \neq J$, we have

$$\frac{\pi_j}{\pi_k} = \frac{F}{1-F}(\eta_j)\frac{1-F}{F}(\eta_k).$$

With designs $Z^{(2)}$ or $Z^{(3)}$, the predictors η_j and η_k are dependent on alternative-specific variables ω_J of the reference alternative. This dependence disappears only if F is the logistic cdf noting that $F/(1-F) = \exp$ and using the exponential property $\exp(\eta_j)/\exp(\eta_k) = \exp(\eta_j - \eta_k)$. If F is not the logistic cdf, then the IIA property holds for the $J-1$ ratios of probabilities π_j/π_J and does not hold for the $\binom{J-1}{2}$ ratios of probabilities π_j/π_k.

2.6. Economic outputs of reference models

Reference models can be used to discuss behavioral hypotheses and elicit economic outputs. To set the theoretical framework, we first discuss the interpretation of the parameter estimates before computing willingness to pay, marginal effects and elasticities formulas. The parameter estimates of reference models can be easily interpreted since

$$\frac{\pi_j}{\pi_J} = \frac{F}{1-F}(\eta_j)$$

is strictly increasing with respect to η_j when $F \in \mathfrak{F}$. For instance, if the slope parameter of the alternatives cost is significantly negative, then the probability of choosing alternative j over alternative J is decreasing when its cost is increasing.

2.6.1. Willingness to pay

For a RUM model, the willingness to pay (WTP) for a continuous alternative specific variable ω (e.g., the time) is the change in cost that keep utilities (deterministic part) unchanged after increasing the ω_{j_0} value by one unit (for a given alternative j_0). It can be shown that no change in utilities implies no change in choice probabilities. The WTP can therefore be defined as the change in cost that keeps the choice probabilities unchanged after increasing ω by one unit.

Property 1. *Assume that there is no interaction between the alternative specific variable ω and the other variables. Then for a reference model the WTP is given by*

$$\mathrm{WTP} = -\frac{\gamma^{\omega}}{\gamma^{\mathrm{cost}}}.$$

In the same way, the value of time (VoT) is defined as the change in cost that keeps the choice probabilities unchanged after increasing the time by one unit. Additionally, the time equivalent comfort (TEC) is the amount of time Δt_{j_0} for a given alternative j_0, necessary to compensate a transition from guarantee of a seating position (comf $= 1$) to no guarantee (comf $= 0$).

Property 2. *Assume that there is an interaction between time and comfort. Then for a reference model the VoT and TEC are given by*

$$\mathrm{VoT} = -\frac{\gamma^{\mathrm{time}} + \gamma^{\mathrm{time}\times\mathrm{comf}}\mathbf{1}_{\mathrm{comf}=1}}{\gamma^{\mathrm{cost}}},$$

$$\mathrm{TEC} = -\frac{\gamma^{\mathrm{comf}} + \gamma^{\mathrm{time}\times\mathrm{comf}} t_{j_0}}{\gamma^{\mathrm{time}}}.$$

Note that the WTP, VoT and TEC formula obtained for reference models are the same as for the MNL. Proofs of Properties 1 and 2 are given in Appendix C.

2.6.2. Marginal effects and elasticities

Marginal effects measure the extent to which probabilities change in response to a change in the variable value. The marginal effect on the probability of choosing the alternative i in the continuous alternative-specific variable ω_j (viewed as a scalar, i.e., only the time value for instance) of alternative j is defined by $\partial\pi_i/\partial\omega_j$ and denoted by $\mathrm{me}_i(\omega_j)$.

Property 3. *For a reference model the marginal effects are given by*

$$\mathrm{me}_i(\omega_j) = \gamma\pi_i \begin{cases} d_j(1-\pi_j) & ,\ i=j,\ j\neq J, \\ -d_j\pi_j & ,\ i\neq j,\ j\neq J, \\ \sum_{k\neq J} d_k\pi_k & ,\ i=j,\ j=J, \\ \sum_{k\neq J} d_k\pi_k - d_i & ,\ i\neq J,\ j=J, \end{cases}$$

where $d_i = f(\eta_i)/[F(\eta_i)\{1-F(\eta_i)\}]$ for $i = 1, \dots, J-1$ and $f = F'$ is the probability density function associated with the cdf F.

The proof is given in Appendix D. Elasticities are an alternative measure of the response of probabilities to a change in the variable value. Their advantage is that they are normalized for the variables' units and are easily computed since

$$\mathrm{e}_i(\omega_j) = \frac{\partial \ln \pi_i}{\partial \ln \omega_j} = \frac{\omega_j}{\pi_i}\mathrm{me}_i(\omega_j), \tag{3}$$

for all $i, j \in \{1, \dots, J\}$. The elasticities are thus given by

$$\mathrm{e}_i(\omega_j) = \gamma\omega_j \begin{cases} d_j(1-\pi_j) & ,\ i=j,\ j\neq J, \\ -d_j\pi_j & ,\ i\neq j,\ j\neq J, \\ \sum_{k\neq J} d_k\pi_k & ,\ i=j,\ j=J, \\ \sum_{k\neq J} d_k\pi_k - d_i & ,\ i\neq J,\ j=J. \end{cases}$$

If $i = j$ then $\mathrm{e}_i(\omega_j)$ is an own-elasticity, whereas if $i \neq j$ then $\mathrm{e}_i(\omega_j)$ is a cross-elasticity. It can be seen that two cross-elasticities $e_i(\omega_j)$ and $e_{i'}(\omega_j)$ are equal when $j \neq J$ but are different when $j = J$. This is a consequence of the IIA property that does not hold for all pairs of alternatives. The change of value in ω_J does not impact the probability of choosing other alternatives in the same way.

Recall that marginal effects and elasticities computation have been presented only for one individual. The aggregated elasticities are computed as follow

$$\mathrm{e}_i(\omega_j) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{e}_i^{\{n\}}(\omega_j),$$

where n denotes the individual index and N the number of individuals. Owing to the IIA property, cross-elasticities have little behavioral value. To get around this issue, Louviere et al. (2000) discussed the method of sample enumeration to compute the weighted elasticities

$$\mathrm{we}_i(\omega_j) = \frac{1}{\sum_{n=1}^{N}\pi_i^{\{n\}}}\sum_{n=1}^{N}\pi_i^{\{n\}}\mathrm{e}_i^{\{n\}}(\omega_j).$$

It should be noted that that the estimation of VoT, TEC, probabilities, marginal effects and elasticities are invariant under the normalization of parameters; see Appendix D.

Table 2
Descriptive statistics.

Variable definition	Label	Mean	S.D.	Min	Max
Alternative-specific variables ($n = 6373$)					
Travel cost by train (in euros)	Cost Train	8.88	7.56	1	62
Travel cost by coach (in euros)	Cost Coach	8.90	7.80	1	78
Travel cost by car (in euros)	Cost Car	9.99	8.64	1	62
Travel time by train (in minutes)	Time Train	69.89	51.86	7	325
Travel time by coach (in minutes)	Time Coach	70.22	53.23	7	325
Travel time by car (in minutes)	Time Car	57.24	36.89	4	330
Comfort in train (1 if seating position guaranteed)	Comfort	0.50			
Individual variables ($n = 1774$)					
Age (in years)	age	47.57	15.76	19	90
Number of cars available per person in the household	access_car	0.76	0.38	0.12	4.00
% of the origin municipality in high density area	Orig_dens	74.21	39.58	0	100
% of the destination municipality in high density area	Dest_dens	86.15	29.93	0	100
Monthly income above 4000 euros (1 is yes, 0 otherwise)	Income_h	0.28			
User of coach for the reference trip (1 is yes, 0 otherwise)	Type_coach	0.03			
User of car for the reference trip (1 is yes, 0 otherwise)	Type_car	0.54			
Makes the reference trip on a regular basis (1 is yes, 0 otherwise)	Regular	0.28			
Imperative schedule at destination (1 is yes, 0 otherwise)	Imperative	0.45			
Car user who had already used public transport to make the reference trip (1 is yes, 0 otherwise)	Alt_pt	0.15			
Train or coach user who had already used car to make the reference trip (1 is yes, 0 otherwise)	Alt_car	0.26			
Frequent use of modes other than car (1 is yes, 0 otherwise)	Freq_alt	0.41			

3. Reference models: an application to mode choice

3.1. Travel stated preferences survey

In this section, we propose an empirical application of the reference models to assess their added-value compared to the MNL model.[3] For this purpose, we use a recent choice experiment on transport mode choice. 1774 inhabitants[4] of the Rhône-Alpes region had to choose between three travel modes[5]: coach ($j = 1$), car ($j = 2$) and train ($j = 3$). Modes vary according to travel time, travel cost and comfort (see Fig. 1 in Supplementary Material). Travel time is defined from origin to destination (including access time, egress time, waiting time and in-vehicle time). Travel cost includes public transport ticket or pass, gasoline, parking cost and toll. Comfort is defined as the guarantee of seating availability. It is therefore always equal to one for coach and car alternatives but sometimes equal to zero for the train alternative, if a seat is not guaranteed. In this case, the train user may have to stand during all or part of the journey. A pivot Bayesian efficient design was implemented (Rose and Bliemer, 2007). A priori weights of attributes were taken from the literature and adjusted during the pilot tests. The percentage of respondents who chose the train alternative was 29.6%, while 21% chose the coach alternative and 49.4% chose the car alternative. Respondents had to answer to four choice questions leading to a database with 6373 observations since a few respondents did not answer all four questions.

Eventually, the alternative-specific variables $(\omega_j)_{j=1,2,3}$ are the travel time, travel cost and comfort as well as the cross-variable Time × Comfort, which is equal to travel time when a seat is guaranteed and equal to zero otherwise. The individual variables x are related to socio-economic characteristics (age and income), spatial characteristics (% of the origin and destination municipalities in a high density area), journey characteristics (train, coach or car user for the reference trip, frequency of the trip, and so on) and general mobility indicators (access to a car and frequency of use of alternative modes to the car). Table 2 displays descriptive statistics for all variables used in the models.

3.2. Specification and design of the reference models

To apply reference models, the first step is to choose the reference alternative. Train is chosen because analyzing train use was the primary objective of the survey. It is therefore the first alternative displayed in the choice questions and may be used as a reference by a majority of respondents. Without *a priori* information about the reference alternative, the modeler could also refer to the goodness-of-fit indicators. For information, the ranking of the models, for each reference alternative, is presented in Supplementary Material; see Table 2.[6]

[3] For further details on this application, including the survey, further interpretation of results and comparison with the literature, see Bouscasse (2017, Chapter 8).

[4] Only respondents aged 18 or over, having a car and a driving license and whose trip was made or could have been made by train or coach were asked to answer the choice questions.

[5] Levels of the attributes were pivoted around the values collected for a reference journey.

[6] The Supplementary Material extends the work carried out in Section 3.3.1 with the alternative train as a reference. It appears that the ranking of cdfs is only marginally modified by the reference alternative. The conclusions of the paper on the performance of the Student cdf and the interest of the reference models compared to the MNL are not questioned.

Table 3
Goodness-of-fit indicators of (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 1, 2, 3$.

(reference, F, $Z^{(1)}$)		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{1.35}$
Indicator value	LL	−5,363	−5,367	−5,356	−5,358	−5,366	−5,357
	$\bar{\rho}^2$	0.230	0.230	0.231	0.231	0.230	0.231
	AIC	10,778	10,785	10,765	10,767	10,784	10,767
	BIC	10,954	10,961	10,941	10,943	10,959	10,950
Rank	LL/$\bar{\rho}^2$/AIC	4	6	1	3	5	2
	BIC	4	6	1	2	5	3
(reference, F, $Z^{(2)}$)		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{0.35}$
Indicator value	LL	−6,083	−6,115	−6,028	−5,966	−6,073	−5,908
	$\bar{\rho}^2$	0.130	0.126	0.138	0.147	0.132	0.155
	AIC	12,178	12,242	12,068	11,945	12,157	11,830
	BIC	12,219	12,282	12,109	11,985	12,198	11,877
Rank	LL/$\bar{\rho}^2$/AIC/BIC	5	6	3	2	4	1
(reference, F, $Z^{(3)}$)		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{1.9}$
Indicator value	LL	−4,803	−4,823	−4,791	−4,797	−4,782	−4,789
	$\bar{\rho}^2$	0.310	0.310	0.311	0.311	0.313	0.312
	AIC	9,665	9,705	9,642	9,654	9,624	9,639
	BIC	9,868	9,908	9,845	9,856	9,826	9,849
Rank	LL/$\bar{\rho}^2$/AIC	5	6	3	4	1	2
	BIC	5	6	2	4	1	3

Notes: LL stands for Loglikelihood.

Reference models are estimated for the three designs $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$. The individual variables were chosen using step-wise selection and doing numerous tests to choose their specification. Using the downgraded designs $Z^{(1)}$ or $Z^{(2)}$ instead of the full design $Z^{(3)}$ implies higher unobserved heterogeneity due to omitted variables. Such models with omitted variables, even if exaggerated for emphasis, allow us to identify the cdfs that are less sensitive to the omission of variables and that retrieve the most consistent economic outputs.

For each design $Z^{(i)}$, the degree of freedom ν is first estimated for the (reference, Student_ν, $Z^{(i)}$) model and then the five other models (using logistic, normal, Laplace, Cauchy and Gumbel cdfs[7]) are estimated. To facilitate comparisons, each triplet (reference, $F^{[r]}$, Z) indicates the rank r of the model according to the $\bar{\rho}^2$ value. For each cdf, the normalized space $\mathfrak{F}_{q_{0.95}}$ has been retained. Indeed, in comparison to the other normalized spaces, it improves the readability of results, the parameter's estimates being less dispersed; see Table 1 for the corresponding values of m_0 and s_0 satisfying Eq. (A.1).

For each design $Z^{(i)}$, the six (reference, F, $Z^{(i)})_{F\in\mathfrak{F}^0}$ models are compared in terms of goodness-of-fit indicators: loglikelihood, $\bar{\rho}^2$ as defined in Ben-Akiva and Swait (1986), AIC and BIC,[8] regarding absolute values as well as ranking. Note that all cdfs have the same number of parameters except the Student distribution, which requires the estimation of only one additional parameter: the degree of freedom ν.

3.3. Interpretation of the results

3.3.1. Goodness-of-fit indicators

The degrees of freedom that maximize the likelihood obtained with (reference, Student_ν, $Z^{(i)})_{i=1,2,3}$ models are respectively $\hat{\nu} = 1.35$, $\hat{\nu} = 0.35$ and $\hat{\nu} = 1.9$; see Appendix B for details about the estimation of ν. With design $Z^{(1)}$ (only individual characteristics), the three best models are the (reference, $F^{[1]} = \text{Laplace}$, $Z^{(1)}$), (reference, $F^{[2]} = \text{Student}_{1.35}$, $Z^{(1)}$) and (reference, $F^{[3]} = \text{Cauchy}$, $Z^{(1)}$) models; see Table 3. According to the Ben-Akiva and Swait test Ben-Akiva and Swait (1986), these three models are not significantly different from one another at the 5% level; see Table 1 in Supplementary Material. However, they outperform the other three models and the (reference, $F^{[5]} = \text{Gumbel}$, $Z^{(1)}$) is not significantly different from the (reference, $F^{[6]} = \text{normal}$, $Z^{(1)}$) model. With design $Z^{(2)}$ (only alternative attributes), the (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) model is the best and (reference, $F^{[2]} = \text{Cauchy}$, $Z^{(2)}$) model follows. All models are significantly different from one another at the 5% level.

With design $Z^{(3)}$ (full model), the (reference, $F^{[1]} = \text{Gumbel}$, $Z^{(3)}$) model is the best and (reference, $F^{[2]} = \text{Student}_{1.9}$, $Z^{(3)}$) and (reference, $F^{[3]} = \text{Laplace}$, $Z^{(3)}$) models follow. As with design $Z^{(2)}$, all models are significantly different from one another at the 5% level.

$Z^{(2)}$ design assumes generic parameters for the alternatives attributes, whereas $Z(1)$ assumes a more flexible form with more parameters associated to individual characteristics. $Z^{(2)}$ models present worse fit than $Z^{(1)}$ models. As expected, $Z^{(3)}$ models show the best fits, as they include all types of variables (individual and alternative-specific).

[7] The (reference, Gompertz, $Z^{(i)}$) models have also been estimated, however results are not presented here because of convergence problems; see Peyhardi (2013, Chapter 5) for theoretical explanations.

[8] The AIC is usually preferred to select a model for prediction whereas the BIC is preferred regarding explanation purposes. Both are retained since both objectives are pursued and they offer similar results.

Table 4
Parameter estimates of (reference, $F_{m,\,s}$, $Z^{(1)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$.

			logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{1.35}$
Alternative-specific constants		Coach	−0.635 *	−0.674 *	−0.492 *	−0.303 **	−0.555 **	−0.412 **
		Car	−0.765 **	−0.791 **	−0.611 **	−0.375 **	−0.554 **	−0.525 **
Individual variables	Coach	Age	0.059	0.052	0.063 *	0.036 *	0.048.	0.047 *
		Income_h	0.302 ***	0.322 ***	0.258 ***	0.125 ***	0.218 ***	0.178 ***
		Orig_dens	0.024	0.027	−0.005	−0.008	−0.021	0.001
		Dest_dens	−0.404 **	−0.418 **	−0.358 ***	−0.174 ***	−0.278 **	−0.243 ***
		Type_coach	2.07 ***	2.232 ***	1.736 ***	0.979 ***	1.8 ***	1.291 ***
		Type_car	1.075 ***	1.193 ***	0.771 ***	0.414 ***	0.867 ***	0.592 ***
		Regular	−0.128	−0.13	−0.111	−0.056	−0.09	−0.081.
		Imperative	−0.072	−0.09	−0.027	0.01	−0.037	−0.005
		Alt_pt	0.021	0.009	0.038	0.05	0.127	0.053
		Alt_car	0.014	0.008	0.012	0.023	0.05	0.026
		Freq_alt	−0.129	−0.138	−0.122.	−0.044	−0.055	−0.066
		Access_car	0.23.	0.232.	0.201 *	0.096.	0.191 *	0.138 *
	Car	Age	0.082 *	0.067.	0.09 **	0.062 **	0.069 *	0.081 **
		Income_h	0.221 **	0.207 *	0.237 **	0.165 **	0.18 **	0.204 **
		Orig_dens	0.145	0.138	0.118	0.069	0.02	0.107
		Dest_dens	−0.286 *	−0.253 *	−0.315 *	−0.219 **	−0.206 *	−0.266 *
		Type_coach	1.789 ***	1.947 ***	1.496 ***	0.842 ***	1.454 ***	1.102 ***
		Type_car	2.796 ***	2.956 ***	2.499 ***	1.605 ***	2.405 ***	1.988 ***
		Regular	−0.662 ***	−0.647 ***	−0.688 ***	−0.522 ***	−0.519 ***	−0.598 ***
		Imperative	−0.321 ***	−0.326 ***	−0.292 ***	−0.167 ***	−0.244 ***	−0.224 ***
		Alt_pt	−0.533 ***	−0.548 ***	−0.516 ***	−0.323 ***	−0.214 *	−0.394 ***
		Alt_car	0.17	0.158	0.193 *	0.142 **	0.098	0.168 *
		Freq_alt	−0.544 ***	−0.557 ***	−0.522 ***	−0.293 ***	−0.387 ***	−0.38 ***
		Access_car	0.415 ***	0.386 **	0.394 ***	0.247 ***	0.405 ***	0.329 ***

*** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level;. = significant at the 10 % level.

Table 5
Parameter estimates of (reference, $F_{m,\,s}$, $Z^{(2)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$.

		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{0.35}$
Alternative-specific constants	Coach	−0.67 ***	−0.739 ***	−0.515 ***	−0.284 ***	−0.562 ***	−0.015 ***
	Car	0.176 ***	0.185 ***	0.138 ***	0.081 ***	0.185 ***	0.004 ***
Alternative-specific variables	Time	−0.012 ***	−0.011 ***	−0.013 ***	−0.01 ***	−0.012 ***	−0.001 ***
	Cost	−0.147 ***	−0.141 ***	−0.135 ***	−0.091 ***	−0.112 ***	−0.006 ***
	Comfort	0.796 ***	0.915 ***	0.554 ***	0.283 ***	0.554 ***	0.013 ***
	Time × Comfort	−0.003 **	−0.004 ***	−0.0008	0.0001	−0.0001	0.00003

*** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level;. = significant at the 10 % level.

Overall, the logistic and normal cdfs perform badly whereas the Student cdf performs well (it is ranked either first or second). Thinner tails' cdfs, like Student or Gumbel seem to fit best with more informative data. With better information, steeper cdf can be better tool to model the change in behavior and can be more precise to estimate switching point.

3.3.2. Interpretation and comparison of parameter estimates

The (reference, F, $Z^{(3)}$) models provide parameter estimates consistent with the theory; see Table 6. For the six models, the time and cost parameters are significant and negative, as expected. The comfort parameter is significant and positive. The crossed effect of time and comfort is positive and significant, indicating the propensity to accept longer travel time in comfortable condition. When significant, the alternative-specific constants indicate that the coach and car alternatives are associated with a smaller baseline probability than the train alternative. Even if the scales of parameters are partly normalized to provide comparable estimations, the direct comparison of parameter estimates is precluded. The focus is therefore only on the structure, sign and significance of the parameter estimates. First, following Viney et al. (2005), Flynn et al. (2008) and Kerr and Sharp (2009), the structure of estimated preferences is analyzed. For each design, the order of parameter estimates is generally respected across models, with relatively stable patterns. Second, when significant, parameter estimates have the same signs across the cdfs for a given design. However, depending on the cdf, parameters differ in their significance level; see Tables 4–6. The choice of the cdf therefore impacts the variables selection. For instance, with design $Z^{(2)}$, the cross-variable Time × Comfort is positive but not significant with the two best performing cdfs ($F^{[1]} = \text{Student}_{0.35}$, $F^{[2]} = \text{Cauchy}$), negative but not significant with the two middle cdfs ($F^{[3]} = \text{Laplace}$, $F^{[4]} = \text{Gumbel}$), and significant and negative with the two worst performing cdfs ($F^{[5]} = \text{logistic}$ and $F^{[6]} = \text{normal}$), which raises a particular concern. The negative sign of the cross-variable Time × Comfort may be due to not capturing the anchoring bias, since the individual reference mode is omitted.

Table 6
Parameter estimates of (reference, $F_{m,s}$, $Z^{(3)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$.

			logistic	normal	Laplace	Cauchy	Gumbel	Student$_{1.9}$
Alternative-specific constants		Coach	−0.54 *	−0.77 **	−0.325	−0.193.	−0.538 **	−0.325.
		Car	−1.137 ***	−1.363 ***	−0.682 **	−0.162	−0.652 ***	−0.597 **
Alternative-specific variables		Time	−0.03 ***	−0.029 ***	−0.028 ***	−0.019 ***	−0.024 ***	−0.027 ***
		Cost	−0.127 ***	−0.123 ***	−0.125 ***	−0.092 ***	−0.098 ***	−0.122 ***
		Comfort	0.46 ***	0.492 ***	0.408 ***	0.23 ***	0.294 ***	0.36 ***
		Time × Comfort	0.003 **	0.003 **	0.003 **	0.002 **	0.003 ***	0.003 **
Individual variables	Coach	Age	−0.009 **	−0.009 **	−0.008 ***	−0.004 **	−0.006 **	−0.007 ***
		Income_h	0.371 ***	0.361 ***	0.316 ***	0.178 ***	0.25 ***	0.297 ***
		Orig_dens	−0.035	0.005	−0.035	−0.031	−0.109	−0.051
		Dest_dens	−0.273 *	−0.181	−0.281 *	−0.166 **	−0.16	−0.265 *
		Type_coach	2.38 ***	2.295 ***	2.237 ***	1.637 ***	2.089 ***	2.225 ***
		Type_car	1.289 ***	1.417 ***	0.97 ***	0.495 ***	1.037 ***	0.911 ***
		Regular	−0.028	0.001	−0.054	−0.037	0.007	−0.044
		Imperative	−0.04	−0.042	−0.01	0.011	−0.021	−0.013
		Alt_pt	0.026	0.011	0.038	0.073	0.137.	0.061
		Alt_car	−0.01	−0.035	0.006	0.013	0.051	0.01
		Freq_alt	−0.213 *	−0.228 *	−0.183 *	−0.061	−0.112	−0.134.
		Access_car	0.138	0.184	0.072	0.006	0.112	0.06
	Car	Age	−0.018 ***	−0.017 ***	−0.02 ***	−0.016 ***	−0.014 ***	−0.019 ***
		Income_h	0.346 ***	0.31 ***	0.38 ***	0.325 ***	0.257 ***	0.377 ***
		Orig_dens	0.15	0.141	0.161	0.116	−0.053	0.146
		Dest_dens	−0.057	0.037	−0.183	−0.261 **	0.019	−0.213
		Type_coach	2.34 ***	2.43 ***	2.186 ***	1.531 ***	1.768 ***	2.084 ***
		Type_car	3.701 ***	3.813 ***	3.35 ***	2.181 ***	2.97 ***	3.135 ***
		Regular	−0.469 ***	−0.408 ***	−0.569 ***	−0.528 ***	−0.37 ***	−0.564 ***
		Imperative	−0.248 **	−0.241 **	−0.235 **	−0.162 **	−0.173 **	−0.238 **
		Alt_pt	−0.276 *	−0.263 *	−0.299 **	−0.193 *	−0.103	−0.272 *
		Alt_car	0.231.	0.254 *	0.191.	0.102	0.141.	0.163
		Freq_alt	−0.599 ***	−0.605 ***	−0.586 ***	−0.33 ***	−0.424 ***	−0.494 ***
		Access_car	0.442 ***	0.443 ***	0.413 ***	0.28 ***	0.415 ***	0.402 ***

*** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level;. = significant at the 10 % level.

Table 7
Value of time (VoT; in € /h.) and time equivalent of comfort (TEC; in minutes) for (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)	logistic	normal	Laplace	Cauchy	Gumbel	Student$_{0.35}$
VoT (comfort = 0)	4.89	4.56	5.82	6.4	6.26	7.42
VoT (comfort = 1)	5.95	6.13	6.16	6.32	6.32	7.10
TEC (time = 30 min)	60.14	75.25	40.51	29.45	47.15	19.03
TEC (time = 60 min)	53.64	64.93	38.76	29.82	46.85	20.32
TEC (time = 90 min)	47.13	54.61	37	30.19	46.54	21.60
(reference, F, $Z^{(3)}$)	**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.9}$**
VoT (comfort = 0)	14.1	14.12	13.56	12.53	14.57	13.39
VoT (comfort = 1)	12.57	12.66	12.24	11.47	12.55	11.98
TEC(time = 30 min)	18.63	20.07	17.37	14.51	16.45	16.37
TEC (time = 60 min)	21.87	23.17	20.28	17.05	20.6	19.53
TEC (time = 90 min)	25.12	26.27	23.18	19.58	24.74	22.68

3.3.3. *Value of time*

Because of the introduction of the cross-variable Time ×Comfort, VoT depends on the guarantee of having a seat; see Property 2. For the (reference, F, $Z^{(3)}$) models, VoT is consistently lower when a seat is guaranteed, since travel time can be used to rest or to work; see Table 7. Depending on the cdf, VoT is comprised between 11.5 and 12.7 € /h when a seat is guaranteed whereas it ranges from 12.5 to 14.6 € /h when a seat is not guaranteed. The ranges of VoT elicited with design $Z^{(3)}$ are in line with literature, which proved to be very heterogeneous depending on mode, travel purpose, type of survey, and so on (Abrantes and Wardman, 2011; Wardman and Wheat, 2013).

The VoT elicited with the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model and (reference, $F^{[5]}$ = logistic, $Z^{(3)}$) models are relatively similar. However, and despite overall consistency of results with design $Z^{(3)}$, some disparities are observed across (reference, F, $Z^{(3)}$) models. They are not due to scale effects since VoT is calculated as a ratio of coefficients. These results highlight the impact of choosing a specific cdf on the estimation of economic outputs, with higher confidence in the results elicited with the best performing cdf.

Table 8
Own-time and own-cost elasticities for (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)	logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{0.35}$
Own-time elasticities						
Train	−0.535	−0.446	−0.696	−0.875	−0.631	−0.979
Coach	−0.601	−0.494	−0.82	−1.062	−0.79	−1.295
Car	−0.281	−0.24	−0.323	−0.424	−0.323	−0.462
Own-cost elasticities						
Train	−0.787	−0.708	−0.854	−0.97	−0.721	−0.948
Coach	−0.875	−0.777	−0.996	−1.169	−0.91	−1.253
Car	−0.583	−0.537	−0.612	−0.653	−0.531	−0.598
(reference, F, $Z^{(3)}$)	logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{1.9}$
Own-time elasticities						
Train	−1.06	−0.998	−1.107	−1.172	−1.128	−1.139
Coach	−1.37	−1.267	−1.47	−1.634	−1.468	−1.534
Car	−0.488	−0.486	−0.465	−0.398	−0.465	−0.453
Own-cost elasticities						
Train	−0.516	−0.488	−0.553	−0.619	−0.507	−0.573
Coach	−0.731	−0.67	−0.821	−0.989	−0.729	−0.866
Car	−0.364	−0.362	−0.355	−0.335	−0.339	−0.357

Whereas the (reference, F, $Z^{(3)}$) models provide consistent results, VoT elicited with the (reference, F, $Z^{(2)}$) models are lower and some cdfs provide inconsistent values; see Table 7. Because of the negative coefficient of the cross-variable Time × Comfort, the VoT is inconsistently lower when a seat is not guaranteed than when a seat is guaranteed, except for the two best models (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) and (reference, $F^{[2]} = \text{Cauchy}$, $Z^{(2)}$). These two models are also the one that elicit the VoT that are the closer to their $Z^{(3)}$ counterparts: (reference, $F^{[2]} = \text{Student}_{1.9}$, $Z^{(3)}$) and (reference, $F^{[4]} = \text{Cauchy}$, $Z^{(3)}$). The Student and Cauchy cdfs therefore seem to be more robust to unobserved heterogeneity than the other cdfs. Overall, the heterogeneity across cdfs is greater with design $Z^{(2)}$ than with design $Z^{(3)}$. VoT range from 4.6 to 7.4 € /h if a seat is not guaranteed and from 6.0 and 7.1 € /h if a seating position is guaranteed. In particular, the best ranking model (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) elicits VoT equal to 7.1 and 7.4 € /h, whereas the VoT elicited with the (reference, $F^{[5]} =$ logistic, $Z^{(2)}$) model are lower and respectively equal to 4.9 and 5.9 € /h.

3.3.4. Time equivalent of comfort

Comfort is converted into minutes to obtain a time equivalent using Property 2; see Table 7. TEC is calculated for three travel times: 30 min, which is the first quartile of the sample; 60 min, the median, and 90 min, the last quartile of the sample. For (reference, F, $Z^{(3)}$) models, TEC increases when travel time increases, as expected. Depending on the cdf, TEC is comprised between 14.5 and 20.1 min of travel time for short travel time, 17.1–23.2 min for medium travel time and 19.6–26.3 min of travel time for long travel time, which is consistent with literature (Richter and Keuchel, 2012; RFF, 2013). As for VoT, TEC are very close between the (reference, $F^{[1]} =$ Gumbel, $Z^{(3)}$) and (reference, $F^{[5]} =$ logistic, $Z^{(3)}$) models, although this does not prevent disparities across (reference, F, $Z^{(3)}$) models.

As expected, the TEC elicited with the (reference, F, $Z^{(2)}$) models are overall inconsistent and range from 19.0 to 75.3 min for short travel time, 20.3–65 min for medium travel time, and 21.6–54.6 min for long travel time. Moreover, TEC inconsistently decrease as travel time increase except for the two best models. The best ranking model is also the only one which elicits TEC in the expected range of value. Indeed, depending on travel time, TEC elicited with the (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) model are comprised between 19.0 and 21.6 min, whereas the (reference, $F^{[5]} =$ logistic, $Z^{(2)}$) model elicits values between 47.1 and 60.1 min. As for VoT, choosing the $\text{Student}_{\nu*}$ distribution over the logistic distribution is therefore, with our data, a means of retrieving consistent outputs.

3.3.5. Elasticities

For own-elasticities, we record only weighted elasticities (Table 8 and Fig..5 in Supplementary Material). For cross-elasticities, unweighted elasticities are recorded to discuss in depth the IIA property and substitution patterns (Table 9). Weighted elasticities are displayed (Table 3) and discussed in Supplementary Material.

Own-elasticities. With design $Z^{(3)}$, the estimated own-elasticities are consistent with those from previous studies that take into account comfort variables (e.g., Daziano and Rizzi, 2015). Also consistent with the literature, the probability of choosing the train or the coach is more sensitive to its own cost and time than the probability of choosing the car (e.g., Glerum et al., 2014). Moreover, a variation of 10% in travel cost has a lower impact on choice probabilities than the same variation in travel time (see De Jong and Gunn, 2001, for similar results).

With design $Z^{(3)}$, the comparison between models offers the same results as for VoT and TEC: own-elasticities are similar between the (reference, $F^{[1]} = \text{Student}_{1.9}$, $Z^{(3)}$) and (reference, $F^{[5]} =$ logistic, $Z^{(3)}$) models, but that does not prevent some disparities between models. With design $Z^{(2)}$, the own-cost elasticities are higher than the own-time elasticities, which seems inconsistent with previous literature (De Jong and Gunn, 2001). The (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) models elicit the own-cost and own-time elasticities closest to each other, and therefore reduce the effect of the misspecification.

Table 9
Unweighted cross-elasticities for (reference, $F_{m,\,s}$, $Z^{(i)}$) models for $F_{m,s} \in \mathfrak{F}_{q_{0.95}}$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{0.35}$
Cross-time elasticities							
Train	Coach	0,252	0,191	0,393	0,589	0,487	0,845
	Car	0,252	0,221	0,28	0,291	0,393	0,198
Coach	Train	0,174	0,139	0,251	0,34	0,218	0,447
	Car	0,174	0,139	0,251	0,34	0,218	0,447
Car	Train	0,31	0,266	0,378	0,446	0,34	0,431
	Coach	0,31	0,266	0,378	0,446	0,34	0,431
Cross-cost elasticities							
Train	Coach	0,355	0,272	0,488	0,69	0,584	0,892
	Car	0,355	0,33	0,343	0,342	0,341	0,212
Coach	Train	0,248	0,215	0,296	0,361	0,245	0,417
	Car	0,248	0,215	0,296	0,361	0,245	0,417
Car	Train	0,637	0,589	0,647	0,679	0,552	0,552
	Coach	0,637	0,589	0,647	0,679	0,552	0,552
(reference, F, $Z^{(3)}$)		logistic	normal	Laplace	Cauchy	Gumbel	$\text{Student}_{1.9}$
Cross-time elasticities							
Train	Coach	0,671	0,455	0,873	1,287	1,201	1,028
	Car	0,671	0,75	0,529	0,26	1,783	0,465
Coach	Train	0,427	0,382	0,474	0,558	0,42	0,502
	Car	0,427	0,382	0,474	0,558	0,42	0,502
Car	Train	0,746	0,801	0,663	0,497	0,631	0,613
	Coach	0,746	0,801	0,663	0,497	0,631	0,613
Cross-cost elasticities							
Train	Coach	0,293	0,158	0,465	0,83	0,553	0,57
	Car	0,293	0,32	0,242	0,139	0,578	0,223
Coach	Train	0,224	0,2	0,257	0,324	0,205	0,275
	Car	0,224	0,2	0,257	0,324	0,205	0,275
Car	Train	0,559	0,605	0,517	0,416	0,46	0,48
	Coach	0,559	0,605	0,517	0,416	0,46	0,48

Cross-elasticities. Any ratio of probabilities of two alternatives i and i' depends on the reference specific-alternative variables ω_J when F is not the logistic cdf. It implies that elasticities $e_i(\omega_J)$ and $e_{i'}(\omega_J)$ are different. For instance, with the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model, when the travel time of train increases by 10%, then the probability of choosing the coach increases by 12.01% and the probability of choosing the car increases by 17.83%. Conversely, a 10% increase in coach travel time induces a 4.2% increase in the probability of choosing the train and coach modes. This higher substitution between the train and car alternatives than between the train and coach alternatives, also verified with the (reference, $F^{[6]}$ = normal, $Z^{(3)}$) model, is contrary to expectations. For the other (reference, F, $Z^{(3)}$) models, as well as the (reference, F, $Z^{(2)}$) models, we consistently observe a higher substitution between public transport modes, except for the worst ranking (reference, $F^{[6]}$ = normal, $Z^{(2)}$) model.

3.3.6. Discussion

Testing a variety of models generates more flexibility in the modeling process and allows the modeler to point out the model that best fits the data. Our results show that the appropriate choice of a cdf and a reference alternative is important to retrieve consistent results, in particular, when the data contain high unobserved heterogeneity since some cdfs are more robust than others. The choice of a distribution and a reference alternative can be guided by *a priori* information and/or goodness-of-fit indicators. The estimation of the model with different cdfs indicates the Student cdf is one of the two best cdfs for all three designs[9] and is the more robust to unobserved heterogeneity. Evaluation of several cdfs for our data indicates that the Student cdf is one of the two best cdfs for all three designs and with all three reference alternatives and is the more robust to unobserved heterogeneity. An explanation is that the flexibility in the degrees of freedom makes it possible to adapt the heaviness of the distribution tails to the data. It seems that the higher the unobserved heterogeneity, the lower the degrees of freedom selected. Indeed, design $Z^{(3)}$ contains the lowest unobserved heterogeneity, and we expect design $Z^{(2)}$ to contain the highest unobserved heterogeneity due to the lack of the individual variables on the type of user (“type_car” and “type_coach”). An interpretation is that thinner tails are needed when data are more informative about behavior. When unobserved heterogeneity is high, the predictor contains little information about the decision maker and the predictor is a poor basis for the detection of changes in choice probability. Distribution tails therefore need to be heavier because the predictor only captures a raw, a gross mean behavior and fails to capture the heterogeneity in behaviors. With heavy-tailed distributions, probabilities are fairly insensitive to changes in predictor values, except around a threshold value around which the slope of the probability function is steep. The Student distribution with low degrees of freedom is able

[9] This result is valid regardless of the reference alternative; see Supplementary Material.

to capture the threshold value at which probability changes are abrupt but cannot capture the more subtle change in probabilities. When unobserved heterogeneity is controlled, a model with an asymmetric distribution stands out. Indeed, with design $Z^{(3)}$, the Gumbel cdf provides good results probably because of its asymmetric shape.

4. Conclusion

Usual economic outputs (VoT, TEC, marginal effects and elasticities) have been theoretically described for the family of reference models. Reference models allow the analyst to derive economic outputs with formulas generalizing those derived with the MNL. For the particular case of the reference model using the logistic cdf, the formulas are those described in the literature; see Appendices C and D.

Estimations of references models on a real dataset describing travel mode choices show that the MNL is outperformed by other reference models. The choice of an appropriate cdf, in particular when the data contain high unobserved heterogeneity.

The heterogeneous results observed between cdfs, and in particular the differences in results between the MNL and the best performing reference models, have operational impacts. First, depending on the chosen cdf, practitioners will not retain the same variables as leverage of public policies. Second, economic outputs, like willingness-to-pay indicators or elasticities, are routinely used in economic models such as cost-benefit analyses or demand forecasting models, which are, in turn, used to derive public policies.

A first avenue for future research is to select the best cdf. On our dataset, the Student cdf is one of the two best cdfs for all three designs and is the more robust to unobserved heterogeneity. Using both individual and alternative specific variables, the Gumbel cdf provides good results because of its asymmetric shape. It would therefore be interesting to test the non-central Student distribution, defined with two parameters: one for the heavy/thin tails and the other for the asymmetry. In order to propose a better comparison of the different reference models, a large simulation study would be appropriate, using cross-validation to select the best cdf. More generally, future research should test reference models on SP, RP and Montecarlo simulated datasets to investigate the performance of the models in different contexts.

A second avenue concerns the choice of the reference alternative. A change of the reference alternative would impact the model fit. Modeler has to test combinations of cdfs and reference alternatives to find the best model fit. Invariance properties under permutations of response categories are studied in Peyhardi et al.. Furthermore, evidence from the economic literature shows that the alternatives chosen as a reference by the decision-makers are certainly not homogeneous across the population (Tversky and Kahneman, 1991). It would thus be relevant to use a model in which different groups of people would have different reference alternatives. One could consider a mixture model with J components corresponding to the J alternatives (in other words, a latent class model). Each individual would be affected to one component, i.e. having one specific alternative as reference. Such an extension is described in Bouscasse (2017), Chapter 8 (Section 8.3.3, p. 279).

Finally, the usual extensions of the MNL, such as the mixed logit model or the nested logit model, can also be developed for reference models. The first extension is straightforward using the GLM framework. The second extension is described by Peyhardi et al. (2016) for the case of nested alternatives, but economic outputs should still be studied.

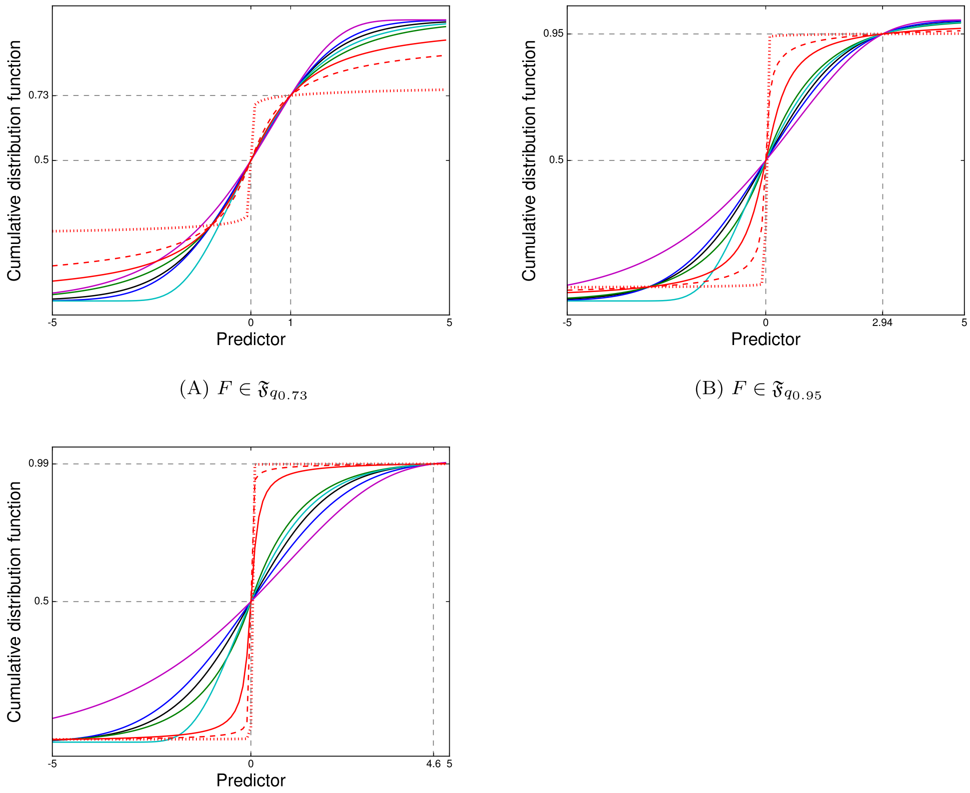
Appendix A. Normalization of parameters via the cdf *F*

The first condition $F(0) = 1/2$, which states that the median is null, is necessary to keep the alternative-specific constant α_j interpretable. Assume that variables x and ω have the same effect on alternative j with respect to the reference alternative J, i.e., $\delta_j = 0$ and $\omega_j = \omega_J$. Therefore, the ratio of probabilities only depends on the alternative-specific constant $\pi_j/\pi_J = F(\alpha_j)/\{1 - F(\alpha_j)\}$ and the null hypothesis $H_0 : \alpha_j = 0$ is thus equivalent to the equality between probabilities π_j and π_J. Testing the significance of the alternative-specific constant α_j is therefore equivalent to testing the significance of the difference between the probabilities π_j and π_J. Note that the condition $F(0) = 1/2$ is already satisfied for symmetric distributions (e.g., logistic, normal, Laplace, Cauchy) and has to be imposed for non-symmetric distributions (e.g., Gumbel and Gompertz).

For the second condition, $F(b) = e^b/(1 + e^b)$, the choice of b is more debatable. Using $b = 1$ seems natural but other values could be used to calibrate the different cdfs F, such as the quantile of the logistic distribution $b = q_p$ for some probability $p \neq 1/2$. For each probability $p \neq 1/2$, we obtain the normalized space $\mathfrak{F}_{q_p} = \{F \in \mathfrak{F} : F(0) = 1/2,\ F(q_p) = p\}$. The logistic, normal, Laplace, Cauchy, Gumbel and Gompertz cdfs $F \in \mathfrak{F}_{q_p}$ are represented in Fig. 2 for the three cases $q_{0.73} = 1,\ q_{0.95} \simeq 2.94$ and $q_{0.99} \simeq 4.6$. In terms of location and scale parameters, we have $F_{m_0,s_0} \in \mathfrak{F}_{q_p}$ if

$$\begin{cases} m_0 = -\dfrac{F^{-1}(1/2) \cdot q_p}{F^{-1}(p) - F^{-1}(1/2)} \\ s_0 = \dfrac{q_p}{F^{-1}(p) - F^{-1}(1/2)}. \end{cases} \tag{A.1}$$

After transformation, the new parameters are $\alpha'_j = (\alpha_j - m_0)/s_0$, $\delta'_j = \delta_j/s_0$ for $j = 1, \ldots, J-1$ and $\gamma' = \gamma/s_0$. The location m_0 implies a translation only of alternative-specific parameters whereas the scale s_0 impacts all the parameters. It should therefore be borne in mind that, even if the parameters are more interpretable after normalization, the best way to compare results obtained with different models is to compare ratios of slopes which stay the same for all normalizations



(A) $F \in \mathfrak{F}_{q_{0.73}}$ (B) $F \in \mathfrak{F}_{q_{0.95}}$ (C) $F \in \mathfrak{F}_{q_{0.99}}$

Fig. 2. Normalized cdf $F \in \mathfrak{F}_{q_p}$ with scales (A) $q_{0.73} = 1$, (B) $q_{0.95} = 2.94$ and (C) $q_{0.99} = 4.6$ using the logistic (black), normal (blue), Laplace (green), Gumbel (cyan), Gompertz (magenta), Cauchy (red), Student$_{\nu=0.5}$ (dashed red line) and Student$_{\nu=0.05}$ (dotted red line) cdfs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(e.g., $\gamma'_{\text{time}}/\gamma'_{\text{cost}} = \gamma_{\text{time}}/\gamma_{\text{cost}}$). Moreover, if the distribution F is symmetric, then $m_0 = 0$ for all normalizations $\left(\mathfrak{F}_{q_p}\right)_{p \neq 1/2}$ and thus the ratios of alternative-specific constants also stay the same (i.e., $\alpha'_j/\alpha'_k = \alpha_j/\alpha_k$ for different alternatives j and k).

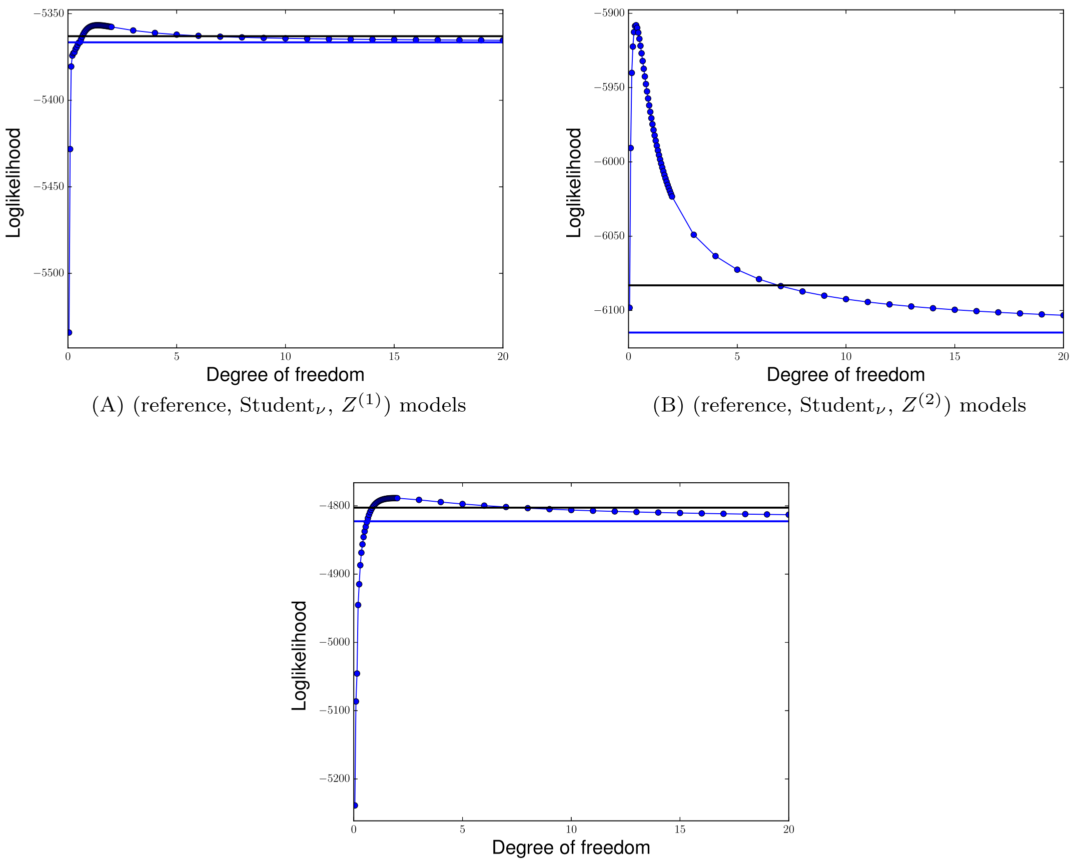
Appendix B. The case of Student distribution

The proposed method is to estimate the (reference, Student$_\nu$, Z) models among a grid of df values and to choose the value $\hat{\nu}$ that maximizes the loglikelihood. We propose to use the following grid: $\nu = 0.05 * i$ for $i = 1, \ldots, 40$ (small value between 0 and 2) and $\nu = 3, \ldots, 20$. The results clearly favor the assumption of loglikelihood concavity; see Fig. 3. Knowing that the Student distribution converges towards the Normal distribution when $\nu \to \infty$, we note that the maximum loglikelihood value obtained with the (reference, Student$_\nu$, $Z^{(i)}$) model converges towards those obtained with the (reference, Normal, $Z^{(i)}$) model when $\nu \to \infty$. It could also be noted that the closer $\hat{\nu}$ is to 0, the higher the gain in loglikelihood obtained with the Student$_{\hat{\nu}}$ distribution in comparison to the normal and logistic distributions; see Fig. 3B and C to compare the case $\hat{\nu} = 0.35$ and $\hat{\nu} = 1.9$. Normal and logistic distribution are unable to capture the abrupt change in probability obtained thanks to the heavy tails of the Student$_\nu$ distribution for small values of ν.

Appendix C. Willingness to pay

C.1. Proof of Property 1

The WTP is the value Δc_{j_0} such that $\pi'_j = \pi_j$ for all alternatives $j = 1, \ldots, J$ after the transformation $\omega'_{j_0} = \omega_{j_0} + 1$ for the specific alternative j_0 (the cost is transformed as $c'_{j_0} = c_{j_0} + \Delta c_{j_0}$). As the link function of a reference model is a surjective map from the simplex to $\mathbb{R}^{J-1}$, for any cdf $F \in \mathfrak{F}$, the equalities between the J probabilities $\pi'_j = \pi_j$ is equivalent to the equalities between the $J-1$ predictors $\eta'_j = \eta_j$. The two cases $j_0 \neq J$ and $j_0 = J$ are presented. If $j_0 \neq J$, only the predictor η'_{j_0}



(A) (reference, Student$_\nu$, $Z^{(1)}$) models

(B) (reference, Student$_\nu$, $Z^{(2)}$) models

(C) (reference, Student$_\nu$, $Z^{(3)}$) models

Fig. 3. Loglikelihood of (reference, Student$_\nu$, $Z^{(i)}$) models for $\nu \in (0, 20]$ (blue points), (reference, normal, $Z^{(i)}$) model (blue horizontal line) and (reference, logistic, $Z^{(i)}$) model (dark horizontal line) on the real dataset with (A) $i = 1$, (B) i=2 and (C) $i = 3$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is impacted by the change $\omega'_{j_0} = \omega_{j_0} + 1$ and we have

$$\begin{aligned} & \eta'_{j_0} = \eta_{j_0} \\ \Leftrightarrow \quad & \gamma^{\omega}\omega'_{j_0} + \gamma^{\text{cost}} c'_{j_0} = \gamma^{\omega}\omega_{j_0} + \gamma^{\text{cost}} c_{j_0} \\ \Leftrightarrow \quad & \gamma^{\omega} + \gamma^{\text{cost}} \Delta c_{j_0} = 0 \end{aligned}$$

If $j_0 = J$ all the $J-1$ predictors are identically impacted by the change ω'_{j_0} and we obtain the same result. Finally, the WTP is given by

$$\text{WTP} = -\frac{\gamma^{\omega}}{\gamma^{\text{cost}}}.$$

C.2. Proof of Property 2

The VoT is the value Δc_{j_0} such that $\pi'_j = \pi_j$ for all alternatives $j = 1, \ldots, J$ after the transformation of the time $t'_{j_0} = t_{j_0} + 1$ for the specific alternative j_0 (the cost is transformed as $c'_{j_0} = c_{j_0} + \Delta c_{j_0}$). As previously, the two cases $j_0 \neq J$ and $j_0 = J$ are presented. If $j_0 \neq J$, only the predictor η'_{j_0} is impacted by the change $t'_{j_0} = t_{j_0} + 1$ and we have (the linear predictor contains the interaction between time and comfort)

$$\begin{aligned} & \eta'_{j_0} = \eta_{j_0} \\ \Leftrightarrow \quad & \gamma^{\text{time}} t'_{j_0} + \gamma^{\text{time}\times\text{comf}} t'_{j_0} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} c'_{j_0} = \gamma^{\text{time}} t_{j_0} + \gamma^{\text{time}\times\text{comf}} t_{j_0} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} c_{j_0} \\ \Leftrightarrow \quad & \gamma^{\text{time}} + \gamma^{\text{time}\times\text{comf}} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} \Delta c_{j_0} = 0 \end{aligned}$$

If $j_0 = J$ all the $J-1$ predictors are identically impacted by the change t'_{j_0} and we obtain the same result. Finally, the VoT is given by

$$\mathrm{VoT} = -\frac{\gamma^{\mathrm{time}} + \gamma^{\mathrm{time}\times\mathrm{comf}}\mathbf{1}_{\mathrm{comf}=1}}{\gamma^{\mathrm{cost}}}.$$

The TEC is obtained using a similar calculation.

Appendix D. Marginal effects calculation

D.1. Proof of Property 3

For the sake of simplicity, we denote by ω_j the value of only one alternative-specific variable, such as the time for instance. The marginal effect of a change in ω_j on the probability of choosing the alternative i is given by

$$\begin{aligned} \mathrm{me}_i(\omega_j) &= \frac{\partial \pi_i}{\partial \omega_j}, \\ &= \frac{\partial \eta}{\partial \omega_j}\frac{\partial \pi_i}{\partial \eta}, \\ \mathrm{me}_i(\omega_j) &= \sum_{k\neq J}\frac{\partial \eta_k}{\partial \omega_j}\frac{\partial \pi_i}{\partial \eta_k}, \end{aligned} \tag{D.2}$$

for all $i \in \{1, \dots, J\}$. On one hand, we have $\eta_k = \alpha_k + x^T\delta_k + (\omega_k - \omega_J)\gamma$ for $k \neq J$ and thus

$$\frac{\partial \eta_k}{\partial \omega_j} = \gamma\left\{\mathbf{1}_{(j=k)} - \mathbf{1}_{(j=J)}\right\},$$

for $k \neq J$ and all $j = 1, \dots, J$. On the other hand, recall that for $i \neq J$ and $j \neq J$

$$\begin{aligned} \frac{\partial \pi_i}{\partial \eta_j} &= d_j \mathrm{Cov}(Y_i, Y_j), \\ \frac{\partial \pi_i}{\partial \eta_j} &= d_j \pi_i\left\{\mathbf{1}_{(i=j)} - \pi_j\right\}, \end{aligned} \tag{D.3}$$

where $d_j = f(\eta_j)/[F(\eta_j)\{1 - F(\eta_j)\}]$ and Y_j is the indicator variables (equal to 1 if j is the chosen alternative and 0 otherwise) for $j \neq J$; see Supplementary Material of Peyhardi et al. for further details. Because of the constraint $\pi_J = 1 - \sum_{i\neq J}\pi_i$, it follows that for $j \neq J$

$$\begin{aligned} \frac{\partial \pi_J}{\partial \eta_j} &= -\sum_{i\neq J}\frac{\partial \pi_i}{\partial \eta_j}, \\ &= -d_j\left\{\pi_j - \pi_j\sum_{i\neq J}\pi_i\right\}, \\ &= -d_j\left\{\pi_j - \pi_j(1-\pi_J)\right\}, \\ \frac{\partial \pi_J}{\partial \eta_j} &= -d_j\pi_j\pi_J, \end{aligned}$$

which means that Eq. (D.3) holds for $j \neq J$ and all $i = 1, \dots, J$. Finally, the equality (D.2) becomes

$$\begin{aligned} \mathrm{me}_i(\omega_j) &= \gamma\left\{\mathbf{1}_{(j\neq J)}\frac{\partial \pi_i}{\partial \eta_j} - \mathbf{1}_{(j=J)}\sum_{k\neq J}\frac{\partial \pi_i}{\partial \eta_k}\right\} \\ \mathrm{me}_i(\omega_j) &= \gamma\pi_i\left[\mathbf{1}_{(j\neq J)}d_j\left\{\mathbf{1}_{(i=j)} - \pi_j\right\} - \mathbf{1}_{(j=J)}\left\{\mathbf{1}_{(i\neq j)}d_i - \sum_{k\neq J}d_k\pi_k\right\}\right] \end{aligned}$$

for all $i = 1, \dots, J$ and all $j = 1, \dots, J$ and thus the desired result is proved.

D.2. The special case of the logistic cdf

The special case of the logistic cdf $F(\eta) = e^\eta/(1+e^\eta)$ is here presented. It should be noted that $f = F(1-F)$ in this case and thus $d_j = 1$ for all $j \neq J$. For alternative-specific variables, we obtain

$$\mathrm{me}_i(\omega_j) = \gamma\pi_i\left[\mathbf{1}_{(j\neq J)}\left\{\mathbf{1}_{(i=j)} - \pi_j\right\} - \mathbf{1}_{(j=J)}\left\{\mathbf{1}_{(i\neq j)} - 1 + \pi_J\right\}\right],$$

and thus $\mathrm{me}_i(\omega_j) = \gamma\pi_i\{\mathbf{1}_{(i=j)} - \pi_j\}$ which is the usual result obtained for the conditional and multinomial logit models (see Greene, 2003, p. 723). The elasticities are given by $\mathrm{e}_i(\omega_j) = \gamma\omega_j\{\mathbf{1}_{(i=j)} - \pi_j\}$. The equality between cross elasticities $\mathrm{e}_i(\omega_j)$ and $\mathrm{e}_{i'}(\omega_j)$ is here true for all $j = 1, \ldots, J$, as expected.

D.3. Invariance of economic outputs under normalization

The VoT, TEC are clearly invariant a normalization of the cdf since they are calculated as ratios of parameters that are only rescaled $\gamma' = s\gamma$. Otherwise, the probabilities π_j stay unchanged since $F_{m,s}(\eta'_j) = F(\eta_j)$ for $j \neq J$ by definition. Now, let us show that marginal effects are not impacted by a normalization even if they are depending on the $d_1, \ldots, d_{J-1}$ quantities; see Property 3. First let us remark that

$$\begin{aligned} f_{m,s}(z) &= \frac{\partial}{\partial z} F_{m,s}(z) \\ &= \frac{\partial}{\partial z} F\left(\frac{z-m}{s}\right) \\ f_{m,s}(z) &= \frac{1}{s} f\left(\frac{z-m}{s}\right) \end{aligned}$$

and thus $f_{m,s}(\eta'_j) = f(\eta_j)/s$, which implies that $d'_j = d_j/s$ for $j \neq J$. After normalization the marginal effects become:

$$\begin{aligned} \mathrm{me}'_i(\omega_j) &= \gamma'\pi_i\left[\mathbf{1}_{(j\neq J)} d'_j \{\mathbf{1}_{(i=j)} - \pi_j\} - \mathbf{1}_{(j=J)}\left\{\mathbf{1}_{(i\neq j)} d'_i - \sum_{k\neq J} d'_k \pi_k\right\}\right], \\ &= \frac{\gamma'}{s}\pi_i\left[\mathbf{1}_{(j\neq J)} d_j \{\mathbf{1}_{(i=j)} - \pi_j\} - \mathbf{1}_{(j=J)}\left\{\mathbf{1}_{(i\neq j)} d_i - \sum_{k\neq J} d_k \pi_k\right\}\right], \\ \mathrm{me}'_i(\omega_j) &= \mathrm{me}_i(\omega_j), \end{aligned}$$

since $\gamma' = s\gamma$, and thus the desired result.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.trb.2018.12.010.

References

Abrantes, P.A., Wardman, M.R., 2011. Meta-analysis of uk values of travel time: an update. Transp. Res. Part A 45 (1), 1–17.
Agresti, A., 2013. Categorical Data Analysis, Third Wiley, New York.
Amemiya, T., 1981. Qualitative response models: a survey. J. Econ. Lit. 19 (4), 1483–1536.
Baltas, G., Doyle, P., 2001. Random utility models in marketing research: a survey. J. Bus. Res. 51 (2), 115–125.
Ben-Akiva, M., Swait, J., 1986. The akaike likelihood ratio index. Transp. Sci. 20 (2), 133–136.
Billot, A., Thisse, J.-F., 1995. Modèles de choix individuels discrets: théorie et applications à la micro-économie. Revue économique 921–931.
Bouscasse, H., 2017. Essays on Travel Mode Choice Modeling - a Discrete Choice Approach to the Interactions Between Economic and Behavioral Theories Ph.D. Thesis.
Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data. comparisons and applications of some estimators and tests. J. Appl. Econom. 1 (1), 29–53.
Daziano, R.A., Rizzi, L.I., 2015. Analyzing the impact of a fatality index on a discrete, interurban mode choice model with latent safety, security, and comfort. Saf. Sci. 78, 11–19.
De Jong, G., Gunn, H., 2001. Recent evidence on car cost and time elasticities of travel demand in europe. J. Transport Econ. Policy 35 (2), 137–160.
De Witte, A., Hollevoet, J., Dobruszkes, F., Hubert, M., Macharis, C., 2013. Linking modal choice to motility: a comprehensive review. Transp. Res. Part A 49, 329–341. doi:10.1016/j.tra.2013.01.009.
Feng, D., Gourieroux, C., Jasiak, J., 2008. The ordered qualitative model for credit rating transitions. J. Empir. Finance 15 (1), 111–130.
Flynn, T.N., Louviere, J.J., Peters, T.J., Coast, J., 2008. Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis. BMC Med. Res. Methodol. 8 (1), 1.
Fullerton, A.S., 2009. A conceptual framework for ordered logistic regression models. Sociol. Methods Res. 38 (2), 306–347.
Glerum, A., Atasoy, B., Bierlaire, M., 2014. Using semi-open questions to integrate perceptions in choice models. J. Choice Modell. 10, 11–33. doi:10.1016/j.jocm.2013.12.001.
Goodman, L.A., 1983. The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. Biometrics 149–160.
Greene, W.H., 2003. Econometric Analysis, Fifth Pearson Education, New York.
Kahn, L.M., Morimune, K., 1979. Unions and employment stability: a sequential logit approach. Int. Econ. Rev. 217–235.
Kerr, G.N., Sharp, B.M., 2009. Efficient design for willingness to pay in choice experiments: evidence from the field. In: Paper Presented at the New Zealand Agricultural and Resource Economics Society Conference.
Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and applications. Cambridge University Press.
Masters, G.N., 1982. A rasch model for partial credit scoring. Psychometrika 47 (2), 149–174.
McCullagh, P., 1980. Regression models for ordinal data. J. R. Stat. Soc. Ser. B (Methodological) 109–142.
McFadden, D., 1973. The measurement of urban travel demand. J. Public Econ. 3 (4), 303–328.
McFadden, D.L., 1978. Modelling the choice of residential location. In: et al., A.K. (Ed.), Spatial Interaction Theory and Residential Location. North Holland, Amsterdam, The Netherlands, pp. 75–96.
McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4 (1), 103–120.

Peyhardi, J., 2013. A New GLM Framework for Analysing Categorical Data; Application to Plant Structure and Development Ph.D. Thesis.
Peyhardi, J., Trottier, C., Guédon, Y.,. A new specification of generalized linear models for categorical responses. Biometrika 102.
Peyhardi, J., Trottier, C., Guédon, Y., 2016. Partitioned conditional generalized linear models for categorical responses. Stat. Modell. 16 (4), 297–321.
RFF, S., 2013. Etude sur la perception et la valorisation de la régularité, du point de vue de l'usager. Technical Report. Réseau Ferré de France.
Richter, C., Keuchel, S., 2012. Modelling mode choice in passenger transport with integrated hierarchical information integration. J. Choice Modell. 5 (1), 1–21.
Rose, J.M., Bliemer, M.C., 2007. Stated preference experimental design strategies. In: Handbook of Transport Modelling: 2nd Edition. Emerald Group Publishing Limited, pp. 151–180.
Sheffi, Y., 1979. Estimating choice probabilities among nested alternatives. Transp. Res. Part B 13 (3), 189–205.
Small, K.A., 1987. A discrete choice model for ordered alternatives. Econometrica 409–424.
Small, K.A., Rosen, H.S., 1981. Applied welfare economics with discrete choice models. Econometrica 105–130.
Tutz, G., 1991. Sequential models in categorical regression. Comput. Stat. Data Anal. 11 (3), 275–295.
Tutz, G., 2012. Regression for Categorical Data. Cambridge University Press, New York.
Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: a reference-dependent model. Q. J. Econ. 106 (4), 1039–1061.
Viney, R., Savage, E., Louviere, J., 2005. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Econ. 14 (4), 349–362.
Wardman, M., Wheat, P., 2013. Meta-Analysis of Post 1994 Values of Non Work Travel Time Savings. Technical Report. University of Leeds.