Journal of Choice Modelling 57 (2025) 100574

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

A smooth bounded choice model: Formulation and application in three large-scale case studies

Laurent Cazor [a,*], Lawrence Christopher Duncan [a], David Paul Watling [b], Otto Anker Nielsen [a], Thomas Kjær Rasmussen [a]

[a] *Technical University of Denmark, Department of Technology, Management and Economics, Akademivej, Building 358, 2800 Kgs. Lyngby, Denmark*
[b] *University of Leeds, Institute for Transport Studies, 36-40 University Road, Leeds LS2 9JT, UK*

ARTICLE INFO

Keywords:
Smooth bounded choice model
One-stage choice set formation
Estimate efficiency
Elasticities
Mode choice
Route choice

ABSTRACT

One-stage (implicit) choice set formation models offer a computationally efficient way to model how individuals consider alternatives. Among these, the Bounded Choice Model (BCM) stands out for its consistent, utility-based cutoffs. However, the BCM is non-differentiable, which limits its usefulness: key outputs such as elasticities and standard errors cannot be computed analytically. To overcome this, we introduce the Smooth Bounded Choice Model (SBCM). This model assumes a new smooth truncated logistic distribution for the error terms and applies a smooth approximation to the maximum function used in defining the reference utility. As a result, the SBCM is infinitely differentiable, while preserving core features of the BCM, such as bounding, continuity, and the ability to collapse to the Multinomial Logit (MNL) model under specific conditions. Importantly, the SBCM is not just a smoother version of the BCM. Its more flexible distributional assumptions can better capture actual choice behaviour and allow for meaningful differences in predicted probabilities. We derive closed-form expressions for choice probabilities, gradients, Hessians, elasticities, and standard errors, and present a practical estimation method. The SBCM is tested in three case studies: one mode choice and two route choice settings (bicycle and public transport). In all cases, it outperforms both the BCM and MNL in terms of model fit and interpretability. While the BCM has so far been limited to car route choice, we show that the SBCM is widely applicable across various discrete choice contexts.

1. Introduction

A crucial step in choice modelling is defining the consideration set from which a decision-maker selects. Modelling consideration set formation is, however, both a computational and behavioural challenge.

The computational challenge in many choice contexts (e.g., in route choice, schedule choice...) comes from the vast number of possible alternatives, which are not feasible to generate and/or operate with. In these cases, the typical modelling approach is to sample from the *universal choice set* (Prato, 2009; Pougala et al., 2021) to identify a *representative universal choice set*. Generating such samples is not trivial: they should include many realistic alternatives and eliminate unrealistic ones while keeping enough variability to capture decision-makers' trade-offs. They must include the chosen alternative for model estimation on revealed preference data. Usually, these choice set generation techniques are not based on behavioural assumptions (Thill, 1992; Prato, 2009; Bovy, 2009), which may lead to biased estimates (Frejinger et al., 2009). Additionally, sampling algorithms may generate unrealistic

* Corresponding author.
E-mail address: lauca@dtu.dk (L. Cazor).

https://doi.org/10.1016/j.jocm.2025.100574
Received 4 July 2024; Received in revised form 16 April 2025; Accepted 29 August 2025
Available online 9 September 2025
1755-5345/

alternatives (Watling et al., 2015), for example, due to the inconsistency between choice set generation and choice probability criteria, which may lead to poor forecasting (Frejinger and Bierlaire, 2010).

The behavioural challenge comes from consideration sets being unobserved. Assuming that individuals choose from the whole (representative) universal choice set may lead to biased substitution patterns (Williams and Ortuzar, 1982). To account for this, Manski (1977) developed a *two-stage* framework where the probability of considering any subset of the representative universal choice set is modelled by a distribution. However, as the number of subsets grows exponentially with the size of the representative universal choice set, this method is not computationally feasible in many real-life choice cases. Many simplifications of this approach have thus been developed (e.g., Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; Tsoleridis et al., 2023), mainly based on heuristics. From the 2000s, a simplified framework combines these two stages by penalising the utility of alternatives that are less likely to be considered based on attribute or utility cutoffs. These models are referred to as *one-stage* or *implicit* choice set formation models (Swait, 2001a; Cascetta and Papola, 2001). Although this framework does not approximate Manski's (Bierlaire et al., 2010), it has the advantage of being computationally tractable. These models have been subject to recent developments and applications in the fields of transportation (e.g., Watling et al., 2018; Yao and Bekhor, 2022; Dubey et al., 2022), land use (e.g., Haque et al., 2019), environmental valuation (e.g., Truong et al., 2015) and marketing (e.g., Swait and Erdem, 2007).

The Bounded Choice Model (BCM, Watling et al., 2018) stands out among one-stage models as the only model that imposes hard, pervasive, compensatory, endogenously defined, and continuous cutoffs on alternatives. What this means and why this is attractive will be discussed in Section 2.2. The BCM belongs to the class of relative random utility models (Zhang et al., 2004) and assumes that individuals do not consider an alternative if it has a deterministic utility much lower than the highest one in the choice set. It does so by assuming a truncated logistic distribution for the utility difference random error terms (rather than a logistic distribution for the multinomial logit model). The difference between the maximum deterministic utility and the truncation threshold is called the *bound* or cutoff: if an alternative deterministic utility is below this bound, it receives zero probability.

However, the BCM has a drawback because its choice probability function is non-differentiable. This property complicates the estimation of parameters, the calculation of standard errors of estimates, and the application of analytical optimisation algorithms. For instance, it is not possible to guarantee the convergence and asymptotic normality of the model parameters' maximum likelihood estimator (Norets, 2010).

To address this issue, this paper introduces the Smooth Bounded Choice Model (SBCM), an infinitely differentiable (or *smooth*[1]) generalisation of the BCM. The SBCM modifies the truncated logistic distribution assumed for the random error terms and uses a smooth approximation of the $\max$ function to ensure the model's smoothness. Crucially, the SBCM maintains the BCM's core properties. Using two extra parameters to control for the choice probabilities smoothness, the SBCM can approximate the BCM to arbitrary precision. The additional contributions of this paper are:

- Derivation of analytical choice probability gradients and Hessian matrices, providing the tools for calculating standard errors of estimates, demand elasticities, and other important metrics.
- Development of a Path-Size corrected SBCM that accounts for overlap in route choice contexts.
- Presentation of an estimation technique that handles the likelihood function's potential non-concavity and constraints on the parameter space.
- Benchmarking and validation through case studies: Comparing the SBCM with the Multinomial Logit (MNL) model and the original BCM in three large-scale case studies in Greater Copenhagen, covering mode choice, bicycle route choice, and public transport route choice. We also study the SBCM elasticities and conduct experiments on the mutual dependency of the smoothing parameters.

The paper is structured as follows. In Section 2, we review the choice modelling literature on the choice set formation problem, motivating our focus on the BCM. In Section 3, we introduce the BCM and demonstrate its non-differentiability. In Section 4, we present the new SBCM and compare it to MNL and the original BCM in illustrative examples. We additionally present an extension of the SBCM to account for route correlation in route choice cases, developing the Smooth Bounded Path-Size model (SBPS). In Section 5, we present a constrained maximum likelihood estimation technique to estimate the model and propose parameterisation techniques to speed up the estimation process. In Section 6, we derive formulas for maximum likelihood estimates' standard errors and demand elasticities. In Section 7, we present the three real-life case studies. In Section 8, we conclude the paper by discussing the results and their implications. We outline other uses of the model and scope for future research.

2. Theoretical background

In this section, we will present the theoretical background for our work. We will begin in Section 2.1 by categorising and discussing the approaches adopted for accounting for choice set formation, highlighting the particular attractiveness of the 'one-stage choice set formation' approach. In Section 2.2, we then review the different models adopting the one-stage choice set formation approach and highlight the attractiveness of the BCM amongst these approaches.

[1] In the following parts of the paper, we will call a function *smooth* if it is infinitely differentiable. In general, a smooth function is a function that is differentiable a sufficient number of times for the function's modelling purposes.

2.1. Modelling choice set formation

Let us denote C the (representative) universal choice set, from which we model the choice probabilities of a decision-maker n. When it comes to modelling choice set formation, approaches can be separated into three main categories:

1. **No choice set formation model:** In the first category, modellers assume individuals choose from the representative universal choice set C, and apply a choice probability model $\mathbb{P}(i|C)$ to determine the probability of choosing each alternative from this set. This method is often employed by real-life applications (e.g., Nielsen et al., 2021)
2. **Two-stage choice set formation models:** In the second category, modellers assume that each decision-maker n chooses from a subset $C_n \subseteq C$ and apply a choice probability model to this subset only ($\mathbb{P}(i|C_n)$). Each individual's consideration set is unobserved and is treated probabilistically. The most prominent approach in this category is the Manski (1977) framework that considers all possible subsets C_n of the representative universal choice set C, and relates the probability of choosing alternative $i \in C$ to $\sum_{C_n \subseteq C} \mathbb{P}(C_n)\mathbb{P}(i|C_n)$ (see, e.g. Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; Başar and Bhat, 2004 for applications and simplifications of this framework).
3. **One-stage choice set formation models:** In the third category, modellers also assume that individuals choose from a subset of the representative universal choice set but that this consideration subset of alternatives C_n is determined implicitly through the computation of the choice probabilities from the choice model (e.g., Cascetta and Papola, 2001; Swait, 2001a; Elrod et al., 2004; Martínez et al., 2009; Paleti, 2015; Truong et al., 2015; Watling et al., 2018 and, more recently, Tan et al., 2024; Kitthamkesorn and Chen, 2024). Rather than modelling the choice set formation with two stages, the consideration stage penalises the utility/probability of alternatives, often based on constraints such as attribute/utility cutoffs.

The first approach assumes that the decision-maker is perfectly rational and has complete information about all available alternatives. This assumption has been widely criticised (starting from Simon (1955)'s work on bounded rationality). Individuals often do not consider all the available alternatives because their number is too high or their relative performance is particularly bad. Not accounting for this leads to misspecification, unrealistic substitution patterns and bias in parameter estimates (Williams and Ortuzar (1982); Frejinger et al. (2009); Bhat (2015)). Ben-Akiva and Lerman (1985) proposes a sampling correction strategy in the case the sampling probability of an alternative from the universal choice set is known (see Frejinger et al. (2009), Flötteröd and Bierlaire (2013) for path choice applications). These methods, however, rely on the property that any alternative from the universal choice set can be selected, which may imply selecting unrealistic alternatives (Frejinger and Bierlaire, 2010) and are thus not suitable for prediction. The sampling error correction strategy also assumes that it is possible to compute the probability of sampling any alternative from the universal choice set, which is not trivial in most applications.

Approaches in the second category allow the consideration choice set to be determined via behaviourally motivated criteria. One would expect the consideration set to be more suitable than the representative set. However, the approach by Manski (1977) is computationally expensive, as the number of subsets grows exponentially with the representative universal choice set size. Thus, most models based on that framework have never been applied in choice contexts where the universal choice set is large, such as route choice modelling. One stage choice set formation models, first theorised by Swait (2001a) and Cascetta and Papola (2001), were initially designed to mimic/approximate Manski's framework. However, Bierlaire et al. (2010) showed that one-stage choice set formation models have different properties and cannot approximate Manski's model, so they should be treated as a standalone modelling framework. One-stage choice set formation models retain the behavioural qualities of the two-stage formation models while keeping a low computational cost. We will focus on this model type in the following subsection and the remainder of the paper.

2.2. One-stage choice set formation models: a review

In this subsection, we explain and present one-stage choice set formation models. Under utility maximisation, if $C_n \subseteq C$ is the choice set of decision-maker n, the choice probability of alternative i is given by:

$$P_n(i) = \Pr(U_{in} \geq U_{jn}, \forall j \in C_n) \tag{1}$$

where $U_{jn} = V_{jn} + \epsilon_{jn}$ is the utility of alternative j for decision-maker n. Assuming the analyst knows C_n, it can be described by deterministic availability indicators (Bierlaire et al., 2010):

$$A_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is considered by individual } n \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The choice model can be re-written:

$$P_n(i) = \Pr(U_{in} + \ln(A_{in}) \geq U_{jn} + \ln(A_{jn}), \forall j \in C) \tag{3}$$

Indeed, if A_{in} is zero, i.e., the alternative i is not available, its $U_{in} + \ln(A_{in})$ will go to minus infinity, and it will have zero probability of being the maximising alternative (Bierlaire et al., 2010). To model for the analyst's lack of knowledge on the actual composition of C_n, Cascetta and Papola (2001) proposed to replace A_{in} by a penalty term $\phi_{in} \in [0, 1]$, which represents the probability that individual n considers alternative i. Generally, alternative penalties depend on their attributes or utilities passing *cutoffs* (also referred to as *bounds* or *thresholds*). Studies from the literature present different functional forms for ϕ_{in}, with different properties (see Appendix A for detailed examples). These properties can be categorised as follows (see Table 1 for examples of references):

Table 1
Key properties of one-stage choice set formation models with references.

Model/Reference	Hard cutoffs	Pervasive cutoff effect	Compensatory cutoff	Endogenous cutoffs	Continuity	Smoothness
Implicit Availability/Perception (IAP), Cascetta and Papola (2001)		×			×	
MNL with cutoff penalties, Swait (2001a)					×	
Generalised nonrectangular hyperbola Elrod et al. (2004)	×	×			×	×
Choice model with screening rules, Gilbride and Allenby (2004)[a]	×		×	×		
Constrained-MNL (C-MNL), Martínez et al. (2009)		×			×	×
C-MNL estimation, Castro et al. (2013)		×		×	×	×
High-order C-MNL, Paleti (2015)		×		×	×	×
Endogenous cutoff model, Truong et al. (2015)				×	×	
BCM, Watling et al. (2018)	×	×	×	×	×	
Conjunctive BCM, Rasmussen et al. (2024)	×	×		×	×	
Truncated Path choice model, Tan et al. (2024)	×	×	×	×	×	
SBCM (this paper)	×	×	×	×	×	×

[a] Gilbride and Allenby (2004) include the possibility for both compensatory and non-compensatory cutoff in their model.

1. **Hard/Soft cutoffs:** A *hard* cutoff means that alternatives with attributes/utilities beyond the cutoff receive zero probability of being considered. *Soft* cutoffs penalise alternatives but never assign them zero probability.
2. **Pervasive/Non-Pervasive cutoffs:** A *pervasive* (as defined by Elrod et al., 2004) cutoff effect means that alternatives which do not invoke the cutoff (i.e., whose utility/attribute value is higher than the lower cutoff value) are still influenced by it. This implies that when an alternative is not cut off by the cutoff, it will still be less likely to be considered if its attribute/utility value is close to the cutoff value than if it is much higher. This is the case for most models in the literature, with Swait (2001a)'s model as a main counter-example.
3. **Compensatory/Non-Compensatory cutoffs:** Cutoffs can be applied to the overall utility (*compensatory*) or using some *non-compensatory* decision rule (e.g., conjunctions/disjunctions of attributes).
4. **Endogenous/Exogenous cutoffs:** Cutoffs are *endogenous* if they depend on attributes of alternatives or can be determined through the estimation of model parameters. *Exogenous* cutoffs are fixed by the analyst independently from attributes of alternatives, or for example, stated by the decision-maker in stated preference data.
5. **Continuous choice probabilities:** The model choice probabilities are continuous with respect to the model parameters and attributes. This is the case of most models in the literature, except for Gilbride and Allenby (2004), which makes the assumption that the choice probabilities are non-continuous at the cutoff value.
6. **Smooth choice probabilities:** Some models have a non-differentiable choice probability function at the cutoff value, which means the choice probabilities have "kinks" when some attribute reaches the cutoff value. Conversely, some models have smooth choice probabilities.

Table 1 displays whether different one-stage choice set formation models have hard or soft cutoffs, pervasive or non-pervasive cutoffs, compensatory or non-compensatory cutoffs, endogenous or exogenous cutoffs, are continuous or non-continuous, and smooth or non-smooth. For each property category, one can argue one side is more attractive than the other.

A hard cutoff can be argued to be more attractive than a soft cutoff as the former assigns strictly zero choice probabilities to alternatives violating cutoffs, and the latter only reduces probabilities. This means that hard cutoffs implicitly generate the consideration set of alternatives from the representative universal choice set, removing unrealistic alternatives as defined consistently by the choice model. This is desirable from a behavioural point of view, especially when the number of potential unrealistic alternatives is vast, e.g., in schedule choice modelling or route choice modelling (Watling et al., 2015). The hard cutoff property is also helpful when comparing "before"/"after" scenarios (e.g., when exploring the implementation of a policy or discount), as the implicit choice set can change and, therefore, adapt with the change in scenario.

A pervasive cutoff can be argued to be more attractive than a non-pervasive cutoff as with the former, the choice probability of an alternative relates to how close it is to the cutoff, i.e. choice probability decreases the closer it gets to the cutoff. This provides clear consistency between choice set formation and calculation of the choice probabilities. It is also supported empirically, where Elrod et al. (2004) found that accounting for the pervasive effect of a cutoff provided a better fit to observed choices.

A compensatory cutoff can be argued to be more attractive than a non-compensatory cutoff, as the former will typically provide consistency between choice set formation and choice probability computation. Most one-stage choice set formation models combine non-compensatory cutoffs for the choice set formation with a compensatory choice from the choice set (e.g. utility maximisation). Horowitz and Louviere (1995) and Swait (2001b) found in an empirical study, however, that the same preferences tend to drive the choice set formation and choice stages, and thus when assuming compensatory choice behaviour, it is attractive to impose compensatory cutoffs. The use of non-compensatory cutoffs is also often linked to a large increase in the number of model parameters, which may make their estimation more complex and lead to identification issues (Castro et al., 2013).

An endogenous cutoff can be argued to be more attractive than an exogenous self-reported cutoff as the latter has been known to cause what has been termed an 'endogeneity issue', where the cutoff is undesirably correlated with the random utility error term (Ding et al., 2012). Respondents often violate their (Moser and Raffaelli, 2014) stated cutoffs, which may lead to poor fit and prediction of models that use them. Truong et al. (2015) found in his study that this led to biased parameter estimates. Endogenous cutoffs are attractive as they can be estimated by estimating the choice model (Duncan et al., 2022) and may depend on the choice situation context. Models with exogenous cutoffs are mainly suitable for stated preference data where respondents state their own cutoff value, like in Swait (2001a).

Lastly, it is attractive for any model to have a continuous and smooth choice probability function. Continuity is, for example, a crucial requirement for estimating the model with MLE (Duncan et al., 2022), and smoothness is an important property for, e.g. evaluating the efficiency of MLE parameter estimates, sensitivity analyses, and efficiently solving equilibrium problems (Castro et al., 2013; Tan et al., 2024).

To summarise, one can argue that it is desirable for a one-stage choice set formation model to have hard, pervasive, compensatory, and endogenous cutoffs, and for the choice probability function to be continuous and smooth. As seen from Table 1, among all the one-stage choice set formation models, a model that satisfies most of these properties is the BCM, which satisfies all properties apart from smoothness. Thus, in this paper, we focus on advancing the BCM, where we resolve the smoothness deficiency.

3. The Bounded Choice Model: formulation and non-differentiability

This section presents the Bounded Choice Model (BCM, Watling et al. (2018)) derivation and highlights some of its properties and non-differentiability.

3.1. Model derivation

The BCM is derived by assuming that each alternative i in the representative or actual universal choice set C is compared to a reference alternative r^* in terms of utility ($U_i = V_i + \epsilon_i, U_{r^*} = V_{r^*} + \epsilon_{r^*}$ where $\epsilon_i, \epsilon_{r^*}$ are random error terms). The distributional assumptions are given for the difference of error terms $\epsilon_{r^*} - \epsilon_i$, which follow a left-truncated Logistic distribution at a threshold $-\phi$, and with a location parameter $\mu = 0$ (see Duncan et al. (2022) supplementary material for a detailed derivation of the model's choice probabilities). Its CDF F_{TL} is given by:

$$F_{TL}(x|\theta,\mu,\phi) = \begin{cases} \dfrac{F_L(x|\theta,\mu) - F_L(-\phi|\theta,\mu)}{1 - F_L(-\phi|\theta,\mu)} & \text{if} \quad x \geq -\phi \\ 0 & \text{if} \quad 0 \leq x < -\phi \end{cases} \tag{4}$$

$$= \frac{(F_L(x|\theta,\mu) - F_L(-\phi|\theta,\mu))_+}{1 - F_L(-\phi|\theta,\mu)} \tag{5}$$

where $(.)_+ = \max(0, .)$. $F_L(x|\theta,\mu) = 1/(1+\exp(-\theta(x-\mu)))$ is the CDF of the Logistic distribution with location parameter μ and scale parameter θ. The BCM choice probability of an alternative i versus the reference alternative is given by:

$$\begin{aligned} \mathbb{P}(\text{choose } i \text{ among } \{i, r^*\}) &= \mathbb{P}(U_i \geq U_{r^*}) \\ &= \mathbb{P}(V_i + \epsilon_i \geq V_{r^*} + \epsilon_{r^*}) \\ &= \mathbb{P}(\epsilon_{r^*} - \epsilon_i \leq V_i - V_{r^*}) \\ &= F_{TL}(V_i - V_{r^*}|\theta, 0, \phi) \end{aligned}$$

The BCM choice probabilities are then given by the ratio of odds ratios (see Watling et al. (2018), Duncan et al. (2022) and Tan et al. (2024)):

$$P_i^{\text{BCM}} := \mathbb{P}(i|C) = \frac{\dfrac{\mathbb{P}(\text{choose } i \text{ among } \{i,r^*\})}{1 - \mathbb{P}(\text{choose } i \text{ among } \{i,r^*\})}}{\sum_{j\in C} \dfrac{\mathbb{P}(\text{choose } j \text{ among } \{j,r^*\})}{1 - \mathbb{P}(\text{choose } j \text{ among } \{j,r^*\})}} \tag{6}$$

$$= \frac{(\exp(\theta(V_i - V_{r^*} + \phi)) - 1)_+}{\sum_{j\in C}(\exp(\theta(V_j - V_r^* + \phi)) - 1)_+} \tag{7}$$

Additionally, Watling et al. (2018) and Duncan et al. (2022) set the reference utility as the maximum deterministic utility, i.e., $V_{r^*} = \max_{j\in C} V_j$.

3.2. Non-differentiability

The non-differentiability of the BCM is twofold:

- The use of a truncated logistic distribution implies a non-differentiability of the choice probabilities around the bound, i.e., when for any alternative $j \in C$, V_j reaches $V_{r^*} - \phi$.
- The definition of the reference alternative $V_{r^*} = V_{j^*}$ where $j^* \in C$ is the index of the deterministic utility maximising alternative implies a non-differentiability with respect to the alternative attributes and utility function parameters.

We detail these two properties below.

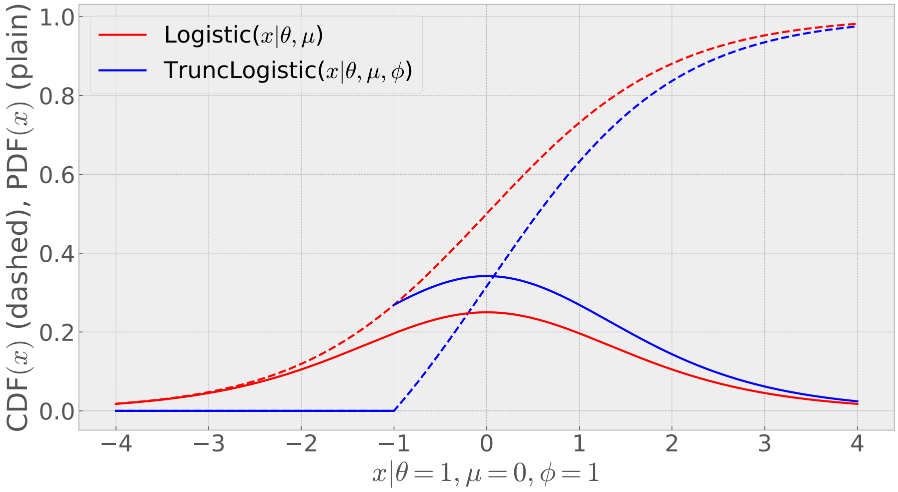


Fig. 1. Logistic and truncated distribution plots, CDF as a dashed line and PDF as a plain line.

Table 2
Example of a choice situation with two alternatives and two attributes.

	TT	TC
Alternative 1	1	2
Alternative 2	2	1

3.2.1. *Non-differentiability around the truncation threshold*

The function $(.)_+$ is non-differentiable at 0, which implies that the CDF F_{TL} is non-differentiable at $-\phi$. The BCM choice probabilities of an alternative versus the reference are thus non-differentiable around the cutoff value ($V_i - V_{r^*} = -\phi$). Consequently, the BCM choice probabilities are also non-differentiable when any of the $V_j, j \in C$ reach the cutoff value ($V_j = V_{r^*} - \phi$). The left-truncated and original logistic distributions are plotted in Fig. 1. We clearly can observe the non-continuity of its PDF and, thus, the non-differentiability of its CDF.

3.2.2. *Non-differentiability of the reference utility*

The $\max$ function is non-differentiable, even though all the V_js are continuous. This non-differentiability happens when the index ($j^*(\boldsymbol{\alpha}) = \arg\max_{j\in C} V_j$) of the utility-maximising alternative changes. Thus, when deterministic utilities are varied, such as taste coefficient parameters being varied during parameter estimation or attributes change in an equilibrium problem, the BCM probability relation in Eq. (7) is non-differentiable. To illustrate this property, let us consider a binary choice situation ($C = \{1, 2\}$). These alternatives have two attributes: travel time (TT) and travel cost (TC). Alternative 1 is fast but expensive; Alternative 2 is cheaper but longer. The attribute values are given in Table 2.

Let us assume that a linear function models the deterministic utility of these alternatives, where for $i \in \{1, 2\}, V_i = \alpha_{TT} TT_i - TC_i$. Using the above notations, we have $\mathbf{X} = \begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}^\top = \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$. Fig. 2 plots how maximum deterministic utility and the derivate of maximum deterministic utility vary as the preference parameter for travel time α_{TT} varies. As can be seen, while the max utility is continuous, the derivate is not, where there is a jump in the derivative at $\alpha_{TT} = -1$. This corresponds to where the reference alternative changes from alternative 1 to alternative 2.

4. The Smooth BCM (SBCM)

In this section, we develop a Smooth Bounded Choice Model (SBCM), which retains the key features of the BCM but has an infinitely differentiable choice probability function, namely by addressing the two smoothness issues we demonstrated in the previous section. An important property of the SBCM is that it can approximate the BCM at any precision and thus can be seen as a generalisation of the model. We first present in Section 4.1 a new bounded support distribution that generalises the truncated logistic distribution, with the additional property that it is smooth for any finite value of its added smoothness parameter. This distribution is then utilised to derive the SBCM choice probabilities, assuming it is the distribution of the error terms difference between any alternative and the reference (Section 4.2). We then present in Section 4.4 an instance of the SBCM, where the reference alternative systematic utility is given by a smooth approximation of the maximum utility in the choice set, whose properties are presented in Section 4.3.

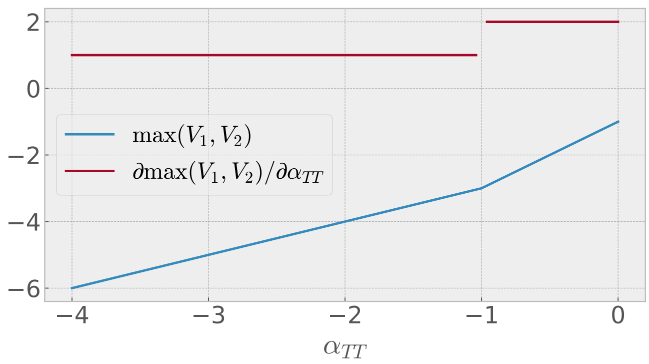


Fig. 2. Evolution of the reference alternative utility in function of α_{TT} on the example from Table 2.

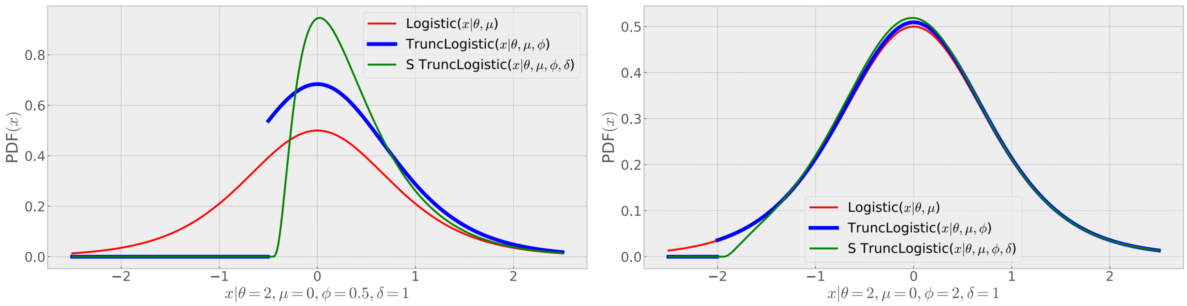


Fig. 3. Plots of the smooth truncated Logistic distribution for $\phi = 0.5$ and $\phi = 3$, the other parameters being fixed.

4.1. A smooth bounded support distribution

As shown in Section 3, the BCM is derived from assuming that the difference in utility between alternative $i \in C$ and a reference alternative r^* follows a truncated logistic distribution. The CDF of the truncated logistic distribution is non-differentiable at the bound, which makes the resulting BCM choice probabilities non-differentiable. To resolve this issue, we propose a smooth variant of this distribution for the random utility difference: a smooth truncated logistic distribution, which has the following CDF:

$$F_S(x|\theta,\phi,\delta) = \frac{g_\delta(\exp(\theta(x+\phi))-1)}{g_\delta(\exp(\theta(x+\phi))-1) + g_\delta(\exp(\theta\phi)+1)}, \tag{8}$$

where $\theta > 0$ is a scale parameter, $\phi > 0$ is a bound parameter, and $\delta > 0$ is the bound smoothing parameter. g_δ is a function given by:

$$g_\delta(z) = \begin{cases} z\exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

As we prove in Appendix B, this distribution has the following properties:

1. It has bounded support on $[-\phi, +\infty)$, $\phi > 0$, so that if X_S follows the distribution, then $\Pr(X_S \le -\phi) = 0$
2. It has a PDF f_S and CDF F_S that are infinitely differentiable on $\mathbb{R}$.
3. It has a *bound smoothing parameter* δ where the distribution converges in distribution to the truncated logistic distribution when δ tends to $+\infty$.
4. **Approximation error**: the approximation error of $(.)_+$ by g_δ is bounded by $1/\delta$. This result implies that the convergence rate of this approximation is controlled by the parameter δ

Note also that for any value of δ, the smooth truncated logistic distribution collapses to the logistic distribution when ϕ tends to $+\infty$, as does the truncated logistic distribution. This is illustrated in Fig. 3, where both the truncated and smooth truncated logistic distributions are plotted with truncation thresholds $\phi = 0.5$ and $\phi = 2$. In Fig. 4, we plot the PDF of the smooth truncated logistic distribution for $\delta = 1$ and $\delta = 15$. One can see that δ has the role of smoothing the PDF around the bound. The larger δ is, the closer the distribution is to the truncated logistic distribution, and the steeper the PDF is around the bound. δ can be estimated as a model parameter through MLE, as we do in Section 7.

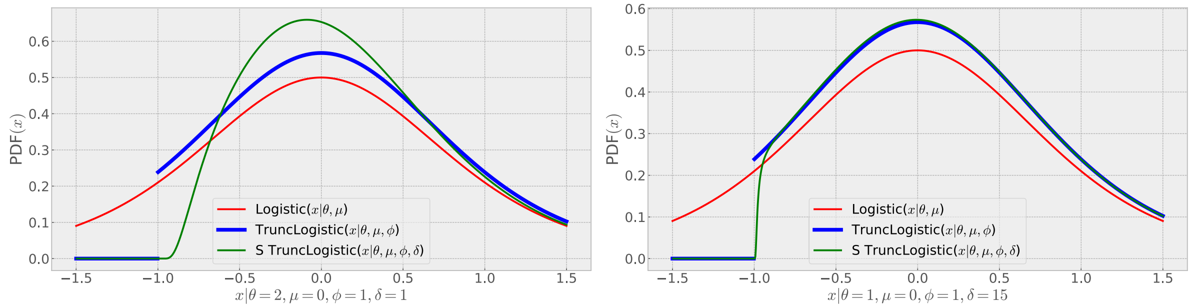


Fig. 4. Plots of the smooth truncated Logistic distribution for $\delta = 1$ and $\delta = 15$, the other parameters being fixed.

From a behavioural standpoint, the assumption of a continuous PDF is appealing, as it avoids implausible discontinuities in the distribution of unobserved preferences. Sharp, deterministic thresholds implied by non-continuous PDFs may not reflect real-world decision-making, where behaviour typically evolves smoothly in response to attribute changes. Nevertheless, the proposed distribution remains flexible enough to accommodate sharply bounded behaviour when appropriate.

4.2. Derivation of the SBCM

Given the smooth truncated logistic distribution introduced in the previous subsection, we shall derive the SBCM probability relation based on this distribution. Assume there is a choice situation with a (representative) universal choice set $C = \{1, \ldots, N\}$. Each alternative $i \in C$ has a random utility $U_i = V_i + \epsilon_i$. Similarly to the derivation made by Watling et al. (2018) for the BCM, we propose that each alternative i is compared with an imaginary reference alternative, whose random utility is $U_{r^*} = V_{r^*} + \epsilon_{r^*}$, in terms of random utility difference. We assume that the random error terms difference $\varepsilon_i = \epsilon_{r^*} - \epsilon_i$, rather than a truncated logistic distribution, follows a smooth truncated logistic distribution at a lower bound $-\phi$, for some $\phi > 0$. Then, the binary probabilities of choosing $i \in C$ over the reference alternative is given by:

$$\begin{aligned}
\mathbb{P}(i|\{i,r^*\}) &= \mathbb{P}(V_i + \epsilon_i \geq V_{r^*} + \epsilon_{r^*}) \\
&= \mathbb{P}(\epsilon_{r^*} - \epsilon_i \leq V_i - V_{r^*}) \\
&= F_S(V_i - V_{r^*}|\theta,\phi,\delta) \\
&= \frac{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1)}{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1) + g_\delta(\exp(\theta\phi) + 1)}
\end{aligned}$$

The odds ratio for alternative $i \in C$ and the reference alternative r^* is then:

$$\begin{aligned}
\eta_i &= \frac{\mathbb{P}(i|\{i,r^*\})}{1 - \mathbb{P}(i|\{i,r^*\})} \\
&= \frac{\dfrac{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1)}{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1) + g_\delta(\exp(\theta\phi) + 1)}}{\dfrac{g_\delta(\exp(\theta\phi) + 1)}{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1) + g_\delta(\exp(\theta\phi) + 1)}} \\
&= \frac{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1)}{g_\delta(\exp(\theta\phi) + 1)}
\end{aligned}$$

The choice probability of alternative i against all the alternatives in the choice set is then given by:

$$P_i^{\mathrm{SBCM},abs} := \mathbb{P}(i|C) = \frac{\eta_i}{\sum_{j\in C} \eta_j} \tag{10}$$

$$= \frac{g_\delta(\exp(\theta(V_i - V_{r^*} + \phi)) - 1)}{\sum_{j\in C} g_\delta(\exp(\theta(V_j - V_r^* + \phi)) - 1)} \tag{11}$$

In addition, Watling et al. (2018) proposed a model for which the bound depends on the reference alternative deterministic utility, i.e., $\phi = (1 - \varphi)V_{r^*}$, with $\varphi > 1$, so that the cutoff is relative to the reference utility. This allows the model to eliminate alternatives from the choice set based on their utility ratio to the reference alternative rather than their utility difference. We shall see in the next subsections that this allows an asymmetry of the probability function and, to some extent, accounts for heteroskedasticity. The choice probability of alternative i is then given by:

$$P_i^{\mathrm{SBCM},rel} := \mathbb{P}(i|C) = \frac{g_\delta(\exp(\theta(V_i - \varphi V_r^*)) - 1)}{\sum_{j\in C} g_\delta(\exp(\theta(V_j - \varphi V_r^*)) - 1)} \tag{12}$$

One important difference between these model formulations is that the absolute model bounds the difference between utilities. In contrast, the relative model bounds the ratio and henceforth needs the utility of every alternative (including the reference alternative) to have the same sign.

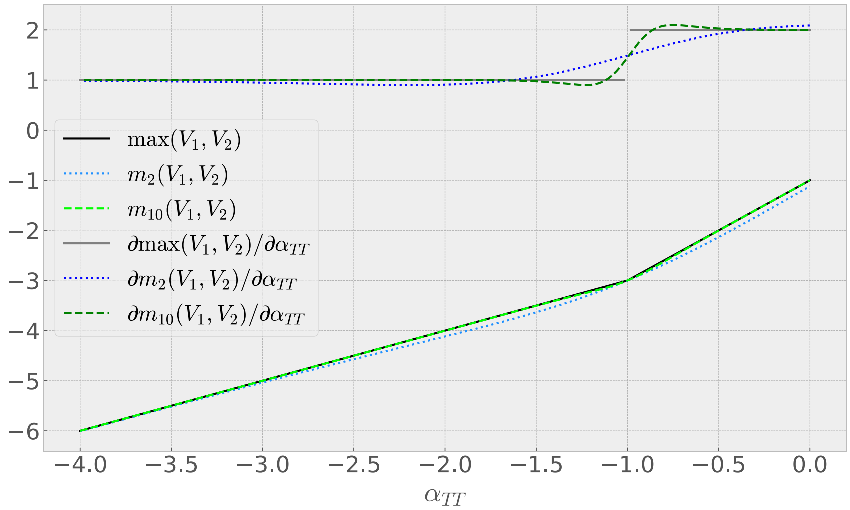


Fig. 5. Maximum and Boltzmann soft maximum operators, and their derivative with respect to α_{TT}: an example for $\lambda = 2, \lambda = 10$.

4.3. Definition of the reference utility

The BCM defines the reference utility as the maximum deterministic utility within the choice set, i.e., for a choice set C, we have that $V_{r^*} = \max_{j \in C} V_j$. As shown in Section 2, since the max function is non-differentiable, so is the BCM choice probability function. Here, we address this issue by approximating the maximum with a smooth approximation. There are numerous approximations for the maximum function available (e.g., the LogSumExp operator, the p-norm, etc.). However, we sought an approximation with the property of never being lower than the actual maximum. This property ensures the SBCM choice probabilities will be defined for any relative bound $\varphi > 1$ and absolute bound $\phi > 0$, which would not be the case if the approximation was smaller than the actual maximum. Indeed, if the reference alternative deterministic utility is larger than the maximum one in the choice set, for some value $\varphi > 1$ or $\phi > 0$, we may get a $0/0$ ratio in the choice probabilities. The approximation we chose was the Boltzmann operator (Asadi and Littman, 2017), which is defined as follows for a vector $\mathbf{x} = (x_1 \cdots x_N)$:

$$m_\lambda(\mathbf{x}) = \frac{\sum_{i=1}^{N} x_i e^{\lambda x_i}}{\sum_{i=1}^{N} e^{\lambda x_i}} \tag{13}$$

where $\lambda > 0$ is a parameter determining the quality of the approximation: as $\lambda \to +\infty$, $m_\lambda(\mathbf{x}) \to \max(\mathbf{x})$. We refer to λ as the *reference utility smoothing parameter*. In determining the maximum deterministic utility, λ also influences the gradient slope when the maximum-utility alternative changes. Considering again the example from Fig. 2 and Table 2, Fig. 5 displays also the Boltzmann operator and its derivative as α_{TT} is varied, with $\lambda = 2$ and $\lambda = 10$. As can be seen, the derivative is continuous around the change of the maximum-utility alternative at $\alpha_{TT} = -1$, where the gradient slope (i.e., the second derivative value) is greater for $\lambda = 10$ than $\lambda = 2$. The Boltzmann operator is also always smaller than the actual maximum.

Behavioural interpretation of m_λ: While the main purpose of the smoothing operator m_λ is to enable differentiability, it has the potential to also capture different behaviours. The smooth maximum can be viewed as the expected value of the vector $\mathbf{x}$ under a softmax weighting scheme:

$$m_\lambda(\mathbf{x}) = \mathbb{E}_p(\mathbf{x}) := \sum_{i=1}^{N} x_i P_i, \quad \text{where} \quad P_i = \frac{e^{\lambda x_i}}{\sum_{j=1}^{N} e^{\lambda x_j}}.$$

From a behavioural standpoint, m_λ can be seen as capturing uncertainty in the decision-makers' perception of the reference alternative's deterministic utility, expressed as $V_{r^*} = m_\lambda(\boldsymbol{V}) + \epsilon_{r^*}$.

At $\lambda = 0$, the operator reduces to the arithmetic mean of all available alternatives, implying that the decision-maker evaluates options relative to an average reference point. In contrast, as $\lambda \to \infty$, the operator converges to the maximum value, suggesting that the decision-maker adopts the highest-utility alternative as the reference point. As will be described in Section 5, λ can be estimated from observed choice data using a maximum likelihood technique.

Approximation error: We prove in Appendix D that the approximation error of the function m_λ to the maximum function decreases with λ by a factor $1/\lambda$.

4.4. Proposed SBCM choice probability relation

Following the derivation in Section 4.2, and replacing the max function with the Boltzmann operator, we propose to use the approximation $V_{r^*} = m_\lambda(\boldsymbol{V})$ where $\mathbf{V} = (V_1, \dots, V_N)$ is the vector of the deterministic utilities in the choice set as the reference

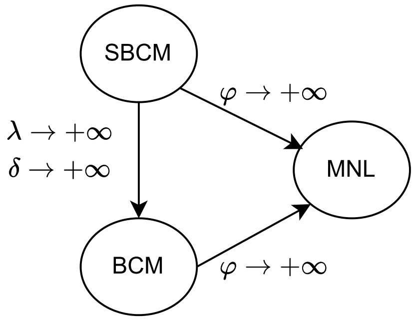


Fig. 6. Model collapsing.

alternative systematic utility. Consequently, the absolute (Eq. (14)) and relative (Eq. (15)) SBCM choice probability relation we propose is as follows for alternative $i \in C$:

$$P_i^{\text{SBCM, abs}} = \frac{g_\delta(\exp(\theta(V_i - m_\lambda(\mathbf{V}) + \phi) - 1))}{\sum_{j \in C} g_\delta(\exp(\theta(V_j - m_\lambda(\mathbf{V}) + \phi) - 1))} \tag{14}$$

$$P_i^{\text{SBCM, rel}} = \frac{g_\delta(\exp(\theta(V_i - \varphi m_\lambda(\mathbf{V})) - 1))}{\sum_{j \in C} g_\delta(\exp(\theta(V_j - \varphi m_\lambda(\mathbf{V})) - 1))} \tag{15}$$

where $\theta > 0$ is the scaling parameter scaling sensitivity to deterministic utility, $\varphi > 0$ is the relative bound parameter determining the cutoff on surplus utility (relative to best utility), g_δ is as in Eq. (9) with $\delta > 0$, and m_λ is as in Eq. (13) with $\lambda > 0$. δ and λ are termed the smoothing parameters, where δ is the bound smoothing parameter and λ is the maximum utility smoothing parameter.

The analytical gradients and Hessian matrix of the SBCM choice probabilities have been derived and are presented in Appendix C.

Since the smooth truncated logistic distribution collapses to the truncated logistic distribution as $\delta \to +\infty$, and the Boltzmann operator collapses to the actual maximum as $\lambda \to +\infty$, the SBCM in Eq. (15) collapses to the BCM in Eq. (7) as $\delta \to +\infty$ and $\lambda \to \infty$. Moreover, both the BCM and SBCM collapse to MNL as $\varphi \to +\infty$ (i.e., there is no screening of alternatives). These collapsing properties are illustrated in Fig. 6. The collapsing properties of the SBCM are theoretically advantageous, as they enable formal statistical comparisons between the SBCM, BCM, and MNL using tools such as likelihood ratio tests. It is important to note, however, that we do not view the SBCM as solely a differentiable approximation of the BCM. Its distributional assumptions extend those of the BCM and may more accurately reflect actual choice behaviour.

Fig. 7 shows the SBCM choice probabilities for $\delta = 1$, compared to the BCM and MNL probabilities with the same relative cost bound (for the BCM) and same scale parameter. The probabilities are plotted on a binary case with utilities (V_1, V_2), and we plot the probability of alternative 2 as a function of V_2, given V_1 is fixed to one. We see that, with this value of δ, the SBCM choice probabilities, while having the same bounding properties as the BCM ones, have a smoother S-shape, which resembles the MNL choice probabilities. This implies that the choice probabilities of alternatives whose utility is close to the bound are much closer to zero for the SBCM than for the BCM.

To illustrate the different properties of the relative and absolute bounding conditions outlined in Section 4, Fig. 8 compares the absolute and relative SBCM choice probabilities for $\delta = 1$, $\varphi = 1.5$ and $\phi = 0.5$, compared to the MNL ones with the same scale parameter. This plot highlights the choice probability asymmetric behaviour. While the two curves overlap for $V_2 > V_1$, the relative bounding leads to a steeper slope of the choice probabilities for $V_2 < V_1$. For the relative model, the utility band on which no alternative is excluded gets larger with larger utility values (e.g., in real life, for longer routes). This is consistent with the heteroskedasticity assumption of asymmetric choice models (e.g., the Multinomial Weibit, Castillo et al. (2008)).

In the remaining parts of the paper, we will mainly focus on the relative version of the SBCM, as it has been the subject of all the further developments of the BCM (Duncan et al., 2022; Tan et al., 2024). This version seems more appealing in some cases, as datasets usually include "small-scale decisions" (e.g., short trips) and "large-scale decisions" (e.g., long trips) so that a relative bounding condition can eliminate alternatives in both these choice situations efficiently, which is the case in the case studies we present in Section 7. However, we do not advocate for choosing a version over another in a general case, as certain cases may be better accommodated with an absolute bound and a symmetric probability function, which does not impose restrictions on the sign of the utility. It is advised to test both versions of the model specifications and to select the one that represents the data best.

4.5. A Smooth Bounded Path Size model capturing route correlations

In this section, we present a route choice extension of the novel SBCM. Due to the complex overlapping nature of road networks, the correlation between routes (i.e. through link-sharing) should be accounted for Florian and Fox (1976) and Cascetta et al. (1996). However, the BCM, and thus the SBCM, do not account for route overlap. Extending the BCM to account for such, Duncan et al. (2022) developed the Bounded Path Size (BPS) route choice model. The BPS model includes heuristic path size correction terms

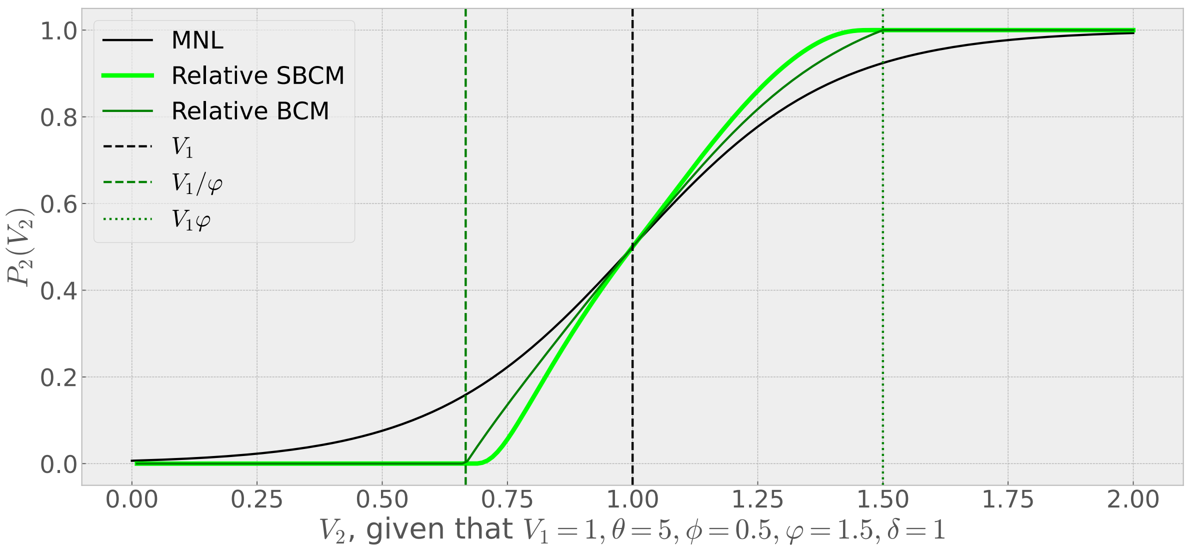


Fig. 7. Comparison of the MNL, relative BCM and SBCM choice probabilities.

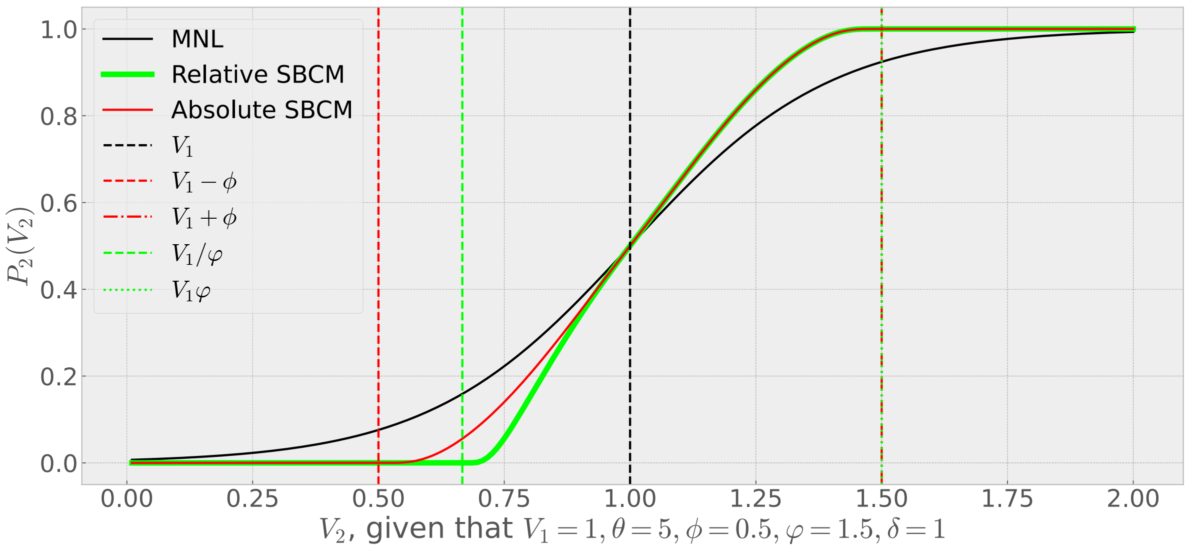


Fig. 8. Comparison of the MNL, Relative and Absolute SBCM choice probabilities.

within the BCM probability relation to adjust the probabilities of used routes when links are shared with other used routes. The model's key feature is that it can capture correlations between only the routes with utilities below the bound, i.e., excluding the impact of routes with utilities above the bound, and do so continuously. For details on how the BPS model is derived and its properties, we direct the reader to Duncan et al. (2022). We shall briefly introduce it and formulate a Smooth BPS (SBPS) model. Define A_i as the set of links constituting route $i \in C$, where link $a \in A_i$ has deterministic utility v_a. The deterministic utility of route i is obtained by summing up the utilities of its constituent links: $V_i = \sum_{a \in A_i} v_a$, where $\mathbf{V}$ is the vector of route utilities. Let $\bar{C} \subseteq C$ be the subset of all routes where $V_i \leq \varphi \max(\mathbf{V})$. The BPS choice probability relation for route $i \in C$ is:

$$P_i^{\mathrm{BPS}} = \begin{cases} \dfrac{\left(\gamma_i^{\mathrm{BPS}}\right)^{\eta} \left(\exp\left(V_i - \varphi \max(\mathbf{V})\right) - 1\right)}{\sum_{j \in \bar{C}} \left(\gamma_j^{\mathrm{BPS}}\right)^{\eta} \left(\exp\left(V_j - \varphi \max(\mathbf{V})\right) - 1\right)} & \text{if } i \in \bar{C} \\ 0 & \text{otherwise} \end{cases}$$

where $(\gamma_i^{\mathrm{BPS}})^{\eta}$ is the path size correction factor for used route $i \in \bar{C}$ (unused routes do not have path size terms). $\eta \geq 0$ is the path size scaling parameter scaling sensitivity to route distinctiveness, and $\gamma_i^{\mathrm{BPS}} \in (0, 1]$ is the path size term for used route $i \in \bar{C}$, calculated as follows:

$$\gamma_i^{\mathrm{BPS}} = \sum_{a \in A_i} \frac{v_a}{V_i} \frac{\left(\exp\left(V_i - \varphi \max(\mathbf{V})\right) - 1\right)}{\sum_{j \in \bar{C}} \left(\exp\left(V_j - \varphi \max(\mathbf{V})\right) - 1\right) \delta_{aj}}$$

where $\delta_{aj} = 1$ if route j uses link a and 0 otherwise. γ_i^{BPS} is specified as such so that (a) unused routes with utilities above the bound, i.e. routes $j \notin \bar{C}$, do not contribute to reducing the path size terms of used routes with utilities below the bound, and (b) the path size term function is continuous as routes enter and exit the used route set $\bar{C}$, as utilities cross from below to above the bound and vice versa. It is also formulated in terms of summing over $j \in \bar{C}$ rather than with $(.)_+$ functions to avoid occurrences of 0/0.

Analogously modifying the BPS model to how we modified the BCM to formulate the SBCM, we also formulate an infinitely differentiable SBPS model:

$$P_i^{\mathrm{SBPS}} = \begin{cases} \dfrac{\left(\gamma_i^{\mathrm{SBPS}}\right)^{\eta} g_\delta(\exp(V_i - \varphi m_\lambda(\mathbf{V})) - 1)}{\sum_{j \in \bar{C}} \left(\gamma_j^{\mathrm{SBPS}}\right)^{\eta} g_\delta(\exp(V_j - \varphi m_\lambda(\mathbf{V})) - 1)} & \text{if } i \in \bar{C} \\ 0 & \text{otherwise} \end{cases}$$

where γ_i^{SBPS} is the Smooth Bounded Path-Size correction term, calculated as follows:

$$\gamma_i^{\mathrm{SBPS}} = \sum_{a \in A_i} \frac{v_a}{V_i} \frac{g_\delta(\exp(V_i - \varphi m_\lambda(\mathbf{V})) - 1)}{\sum_{j \in \bar{C}} g_\delta(\exp(V_j - \varphi m_\lambda(\mathbf{V})) - 1)\delta_{aj}} \tag{16}$$

where g_δ is given by Eq. (9). This weight formulation allows choice probabilities to remain infinitely differentiable when adding the Path-Size correction.

5. Parameter estimation approach

In this section, we discuss the estimation of the SBCM parameters from observed choice data. The different parameters to be estimated can be stored in a vector $\boldsymbol{\beta} = (\boldsymbol{\alpha}, \theta, \varphi, \delta, \lambda) \in \mathbb{R}^{K+3}$, where K is the number of attributes included in the utility function. $\boldsymbol{\alpha} \in \mathbb{R}_+^{K-1}$ are the normalised cost function parameters, θ is the scale parameter, which should be constrained as positive if utility maximisation is the expected behaviour, or negative if cost minimisation is expected. $\varphi > 1$ is the relative utility/cost bound parameter, $\delta > 0$ is the bound smoothing parameter and $\lambda > 0$ is the maximum utility smoothing parameter.

5.1. Estimation technique

Let us assume that we observe N choices. An observation $n \in \{1, .., N\}$ has a choice set C_n, and the index of the chosen alternative is given by $i_n \in C_n$. To estimate the SBCM model and its special cases, we adopt the modified Maximum Likelihood Estimation (MLE) procedure originally proposed in Duncan et al. (2022) for the Bounded Path-Size model. Since for certain specifications of the parameters, a chosen route under the SBCM may receive a zero choice probability, the likelihood function can be zero. As discussed in Duncan et al. (2022) though, the optimal parameters will always lie in the parameter subspace, leading to a non-zero likelihood. If $\boldsymbol{\alpha}$ parameterises the utility functions $V_i(\boldsymbol{\alpha})$, the valid parameter subspace for observation n is given by:

$$\Theta_n = \left\{ (\boldsymbol{\alpha}, \varphi, \lambda), V_{i_n} \geq \varphi^{-1} m_\lambda(\mathbf{V}) \right\}$$

If the model parameters belong to this space, it means that the observed chosen alternative is given a non-zero choice probability (i.e. is not cut-off by the relative utility bound). The valid parameter subspace for the likelihood function is given by the intersection of the valid subspaces for all the observations, i.e.:

$$\Theta = \bigcap_{n=1}^{N} \Theta_n = \left\{ (\boldsymbol{\alpha}, \varphi, \lambda), \forall n \in \{1, .., N\}, V_{i_n} \geq \varphi^{-1} m_\lambda(\mathbf{V}) \right\}$$

This ensures that the model cannot assign zero probability to any chosen alternative. Continuity of the likelihood function is guaranteed over the constrained parameter subspace. To ensure during maximum likelihood estimation that the estimated parameters remain in Θ, we defined the following log-likelihood function. If $\boldsymbol{\beta}$ is the vector containing all the model parameters:

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^{N} \log P_{i_n}^{\mathrm{SBCM}}(\boldsymbol{\beta}) \cdot \mathbb{1}_{(\boldsymbol{\alpha}, \varphi, \lambda) \in \Theta_n} - C \cdot \mathbb{1}_{(\boldsymbol{\alpha}, \varphi, \lambda) \notin \Theta_n} \tag{17}$$

where C is a large penalising constant (in the following case studies, we take $C = 999$) that ensures that the optimum cannot be found outside Θ. $\mathbb{1}_{(\boldsymbol{\alpha}, \varphi, \lambda) \in \Theta_n}$ is the indicator function, that is 1 if and only if the set of parameters is in the subspace Θ_n. This formulation ensures the tuple $(\boldsymbol{\alpha}, \varphi, \lambda)$ remains in the domain Θ when using maximisation algorithms. In this paper, we optimise the log-likelihood using the L-BFGS-B algorithm, using the Python programming language along with the NumPy and SciPy packages. A potential issue is that the SBCM log-likelihood function is not guaranteed to be concave. MLE solutions are thus not guaranteed to be unique according to standard proofs. That is not to say, though, that MLE solutions are or cannot be unique. In Section 7, we explore the uniqueness of MLE solutions numerically by re-conducting MLE with several different randomly generated initial conditions. The solutions found are always the same, suggesting uniqueness, and mirroring similar findings in Duncan et al. (2022) for the Bounded Path-Size model.

5.2. Initialisation

When estimating the model parameters, it is important to make sure that the initial conditions are within Θ, as they may otherwise remain stuck outside Θ. To do so, we used random initialisation for all the model parameters other than the bound, drawing from a standard normal distribution. Then, we calculated the minimum value of the bound so that the utility parameters belong to Θ:

$$\varphi_{\min} = \max_{n\in\{1,\dots,N\}} \frac{c_{i_n}}{\min_{j\in C_n} c_j}$$

Then, we draw a random value of φ_0 from a left-truncated random distribution on $(\varphi_{\min}, +\infty)$. In our paper, we chose to use a shifted exponential distribution with scale parameter 1: $\varphi_0 \sim \mathrm{Exp}(1) + \varphi_{\min}$.

5.3. Reparameterisation techniques

It is well-known that a constrained optimisation problem is more complex to solve than an unconstrained one and requires more advanced and, usually, less efficient algorithms (Nocedal and Wright, 2006). However, the current model formulation imposes constraints on parameter values. To avoid issues with parameter constraints in the log-likelihood estimation, we parameterise the model following the work of Lipovetsky (2009). This transformation allows parameters to remain unconstrained in estimation while ensuring they stay within a reasonable range. We define the reparametrisations as follows:

$$\theta = \exp(\tilde{\theta}) \tag{18}$$

$$\varphi = 1 + \exp(\tilde{\varphi}) \tag{19}$$

$$\lambda = \exp(\tilde{\lambda}) \tag{20}$$

$$\delta = \exp(\tilde{\delta}) \tag{21}$$

where $\tilde{\varphi}, \tilde{\theta}, \tilde{\lambda}$ and $\tilde{\delta}$ are the unconstrained parameters. These transformations prevent δ and λ from reaching arbitrarily large values, improving numerical stability. Similarly, constraints on utility or cost function parameters can be enforced by setting $\boldsymbol{\alpha} = \exp(\tilde{\boldsymbol{\alpha}})$ which ensures all elements of $\boldsymbol{\alpha}$ remain positive or negative, depending on the needed sign restriction. With these transformations, the MLE procedure estimates parameters in the unconstrained space: $\tilde{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{\alpha}}, \tilde{\theta}, \tilde{\varphi}, \tilde{\delta}, \tilde{\lambda}\right) = \boldsymbol{\Phi}^{-1}(\boldsymbol{\beta})$, where the transformation function is defined as

$$\boldsymbol{\Phi}(\boldsymbol{\beta}) = (\exp(\boldsymbol{\alpha}), \exp(\theta), 1 + \exp(\varphi), \exp(\delta), \exp(\lambda)) \tag{22}$$

5.4. Parameter interpretation and significance tests

Interpreting estimated parameters requires applying the inverse transformation to recover $\boldsymbol{\beta}$ from $\tilde{\boldsymbol{\beta}}$. Similarly, standard errors must be adjusted accordingly. Following Daly et al. (2012), the covariance of the transformed parameters is given by:

$$\mathrm{Cov}(\boldsymbol{\beta}) = \mathrm{Cov}(\boldsymbol{\Phi}(\tilde{\boldsymbol{\beta}})) = \boldsymbol{\Phi}'(\tilde{\boldsymbol{\beta}})^{\top}\mathrm{Cov}(\tilde{\boldsymbol{\beta}})\boldsymbol{\Phi}'(\tilde{\boldsymbol{\beta}})$$

where $\mathrm{Cov}(\tilde{\boldsymbol{\beta}})$ is the Asymptotic Variance-Covariance matrix of the estimated parameter $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Phi}'$ is the Jacobian matrix of $\boldsymbol{\Phi}$. Given the definition in Eq. (22), the Jacobian simplifies to

$$\boldsymbol{\Phi}'(\tilde{\boldsymbol{\beta}}) = (\exp(\boldsymbol{\alpha}), \exp(\theta), \exp(\varphi), \exp(\delta), \exp(\lambda))$$

For two parameters $\beta_i, \beta_j \in \boldsymbol{\beta}$, their covariance follows

$$\mathrm{Cov}(\beta_i, \beta_j) = e^{\tilde{\beta}_i} e^{\tilde{\beta}_j}\,\mathrm{Cov}(\tilde{\beta}_i, \tilde{\beta}_j)$$

When conducting significance tests, it is important to note that while utility function parameters and scale parameters are typically compared to zero, the parameters φ, δ and λ should be statistically compared to $+\infty$. This can be done by comparing their inverse to zero. Using Daly et al. (2012), the variance of an inverse parameter is given by

$$\mathrm{Cov}(1/\beta_i, 1/\beta_j) = \frac{\mathrm{Cov}(\beta_i, \beta_j)}{\beta_i^2 \beta_j^2} \tag{23}$$

Applying this to the t-test, we observe that the test statistic for comparing a parameter to zero is equivalent to the test against infinity: $\frac{1/\beta}{\sqrt{\mathrm{Var}(1/\beta)}} = \frac{1/\beta}{1/\beta^2\sqrt{\mathrm{Var}(\beta)}} = \frac{\beta}{\sqrt{\mathrm{Var}(\beta)}}$

6. Advantages of the SBCM over the BCM

In this section, we detail some advantages of SBCM smoothness. The list is non-exhaustive and depends on what the model is used for. Two major advantages of the smoothness property are the ability to analyse the efficiency of estimates analytically and the elasticities of demand to attributes.

We can derive the likelihood function gradients and Hessian matrix of a MLE on observed choices. Suppose we observe N choice situations for which the choice probability of the chosen alternative is given by P_n. In that case, we define the log-likelihood function as $\mathcal{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \ln P_n(\boldsymbol{\beta})$. By linearity of differentiation, we then have:

$$\nabla_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \nabla_{\boldsymbol{\beta}} \ln P_n(\boldsymbol{\beta})$$

$$\nabla^2_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) = \sum_{n=1}^{N} \nabla^2_{\boldsymbol{\beta}} \ln P_n(\boldsymbol{\beta})$$

6.1. Efficiency of estimators

First, one of the regularity conditions for the asymptotic normality of ML estimates is that the Likelihood function is differentiable. Consequently, the non-differentiable property of the standard BCM means that asymptotic normality is not guaranteed. The SBCM, on the other hand, is smooth, and asymptotic normality is guaranteed.

Second, when performing MLE on a sample of observations, one desired output is the efficiency of these estimates. The Cramer–Rao theorem (Harald Cramer, 1946; Radhakrishna Rao, 1945) gives a lower bound for the variance–covariance matrix of the true model parameters. It states if $\boldsymbol{\beta}_t$ are the true parameters and if the model is correctly specified, then the asymptotic variance of the MLE estimated parameters $\hat{\boldsymbol{\beta}}$ is given by:

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) = -\mathbb{E}\left[\nabla^2_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}_t)\right]^{-1} &\approx -\mathbb{E}\left[\nabla^2_{\boldsymbol{\beta}} \mathcal{L}(\hat{\boldsymbol{\beta}})\right]^{-1} \\ &= \left(\sum_{n=1}^{N} \nabla^2_{\boldsymbol{\beta}} \mathbb{E}\left[\ln P_n(\hat{\boldsymbol{\beta}})\right]\right)^{-1} \\ &= \left(\sum_{n=1}^{N} \nabla^2_{\boldsymbol{\beta}} \ln P_n(\hat{\boldsymbol{\beta}})\right)^{-1}\end{aligned}$$

A MLE estimator attains this lower bound if the sample size N tends to infinity. Thus, the new SBCM allows computing the Cramer–Rao bound at the model estimates, i.e., to get their asymptotic variance–covariance (AVC) matrix and their standard errors, on which statistical tests can be performed. Usually, the t-test is performed to assess the significance of the maximum likelihood estimate. For an estimated parameter $\hat{\beta}$, the t-statistic t-stat($\hat{\beta}$) is given by $\hat{\beta}/se(\hat{\beta})$, where $se(\hat{\beta}) = \sqrt{\mathrm{Var}(\hat{\beta})}$. These t-statistics are compared to the critical values (often 95%) of the $T(N_{obs}, 0.975)$, where T is the t-distribution inverse CDF, N_{obs} is the number of observations used for estimation (for any model with more than 1000 observations, this value is similar to the one of the normal distribution, i.e., 1.96).

6.2. Elasticities

Compiling the probability gradients also allows for the computing of elasticities. Elasticities measure the sensitivity of one quantity to another (when both quantities are dependent). Choice modellers often calculate the elasticities of the choice probabilities of an alternative i from a choice set C to one attribute of this alternative (or another). These elasticities output the responsiveness of demand or market shares to a change in one attribute. Elasticities are defined as the marginal change of an alternative's choice probabilities as a function of the marginal change of an attribute of this alternative (or of another alternative for cross elasticities). The disaggregate direct point elasticities can be calculated as:

$$E^{P_i}_{x_{ik}} = \frac{\partial P_i}{\partial x_{ik}} \frac{x_{ik}}{P_i} = \frac{\partial \ln P_i}{\partial x_{ik}} x_{ik} \tag{24}$$

where P_i is the probability of alternative $i \in C$ and x_{ik} is its kth attribute. Similarly, disaggregate cross point elasticities are calculated as follows:

$$E^{P_i}_{x_{jk}} = \frac{\partial P_i}{\partial x_{jk}} \frac{x_{jk}}{P_i} \tag{25}$$

where $j \in C$ is another alternative. Let us define the function f_i, for any $i \in C$, as:

$$f_i(\mathbf{X}|\theta, \varphi, \lambda) = \exp(\theta^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\theta)) - 1 \tag{26}$$

We have:

$$\frac{\partial \ln P_i}{\partial x_{ik}} = \frac{1}{g_\delta(f_i)} \frac{\partial g_\delta(f_i)}{\partial x_{ik}} - \frac{1}{\sum_{j\in C} g_\delta(f_j)} \sum_{j\in C} \frac{\partial g_\delta(f_j)}{\partial x_{ik}}$$

where the partial derivative of the probability numerator to the attribute x_{ik} can be computed with the following chain rule.

$$\frac{\partial g_\delta(f_i)}{\partial x_{ik}} = \frac{\partial g_\delta}{\partial f_i}\frac{\partial f_i}{\partial x_{ik}} \tag{27}$$

with $\partial g_\delta/\partial f_i$ being given by Eq. (52) of Appendix C,

$$\frac{\partial f_i}{\partial x_{ik}} = \theta_k - \varphi\frac{m_\lambda}{\partial x_{ik}} \tag{28}$$

$$\frac{\partial m_\lambda}{\partial x_{ik}} = \theta_k \frac{\exp(\lambda\boldsymbol{\theta}\mathbf{x}_i)}{\sum_j \exp(\lambda\boldsymbol{\theta}\mathbf{x}_j)}\left(1 + \lambda(\boldsymbol{\theta}\mathbf{x}_i - m_\lambda(\mathbf{X}|\boldsymbol{\theta}))\right) \tag{29}$$

The elasticities can then be computed by combining the equations above. For a series of observed choices, $n \in \{1,\dots,N\}$, we call $P_n(i)$ the probability of alternative $i \in \mathcal{C}_n$ for choice situation n. The predicted share of alternative i, which we will call $S(i)$, is given by its average choice probability, i.e., $S(i) = \frac{1}{N}\sum_{n=1}^{N} P_n(i)$. We define the aggregate point elasticity of alternative i with respect to attribute k as the share elasticity with respect to this attribute $E_{x_{ik}}^{S(i)}$:

$$E_{x_{ik}}^{S(i)} = \frac{\partial S(i)}{\partial x_{ik}}\frac{x_{ik}}{S(i)} \tag{30}$$

$$= \sum_{n=1}^{N} E_{x_{nik}}^{P_n(i)} \frac{P_n(i)}{\sum_{p=1}^{N} P_p(i)} \tag{31}$$

This share elasticity represents how much the predicted share of alternative i within the observed choice situations will change if the value of the attribute k changes.

7. Case studies

In this section, we use MLE to estimate the SBCM in three transport case studies: a mode choice case (Section 7.1), a bicycle route choice case (Section 7.2), and a public transport mode/route choice case (Section 7.3). For all case studies, we shall compare the SBCM to the standard BCM and MNL models regarding goodness-of-fit to the data, interpretation of the results, and evaluate estimate efficiency. For the mode choice case study, we shall also analyse the estimation of the smoothing parameters of the SBCM and evaluate aggregate elasticities of parameters. For the bicycle route choice case study, we extend the SBCM to account for route overlap smoothly and investigate how this affects estimation results.

It is worth noting that, in all case studies, the SBCM specification with an absolute bound was also estimated. However, it was consistently outperformed by the relative bound specification and was therefore excluded from the reported estimation results. Two main factors may explain this superior performance. First, the relative bound is more effective in mode and route choice contexts, as it excludes alternatives based on utility ratios rather than absolute differences, making it better suited to both short and long origin–destination movements. Second, the relative specification incorporates asymmetric choice probabilities, which often better reflect actual choice behaviour due to factors such as loss aversion (e.g., Chikaraishi and Nakayama, 2016; Fosgerau and Bierlaire, 2009; Brathwaite and Walker, 2018).

7.1. Mode choice in the Greater Copenhagen Area

The first case study is a mode choice model in the Greater Copenhagen Area. The dataset has been extracted from the Danish National Travel Survey and contains 21,270 mode choice observations collected between 2009 and 2019. The universal choice set of mode alternatives was assumed to be the car, Public Transport (PT), cycling, and walking. However, as we explore through estimation of the SBCM, some of these alternatives are unused in some situations.

7.1.1. Model specification

The estimated models use the following utility specifications, inspired from the Başaran et al. (2025) case study:

$$V_{car} = \alpha_{\mathrm{GTT},car} \times \mathrm{GTT}_{car}$$
$$V_{PT} = \mathrm{ASC}_{PT} + \alpha_{\mathrm{GTT},PT} \times \mathrm{GTT}_{PT} + \alpha_{acc} \times \mathrm{Acc} + \alpha_{egr} \times \mathrm{Egr}$$
$$V_{cycle} = \mathrm{ASC}_{cycle} + \alpha_{\mathrm{GTT},cycle} \times \mathrm{GTT}_{cycle}$$
$$V_{walk} = \mathrm{ASC}_{walk} + \alpha_{\mathrm{GTT},walk} \times \mathrm{GTT}_{walk}$$

ASC are the Alternative Specific Constants, Acc and Egr are the Access and Egress times to the public transport stops, calculated using the methodology from Anderson (2013). The Generalised Travel Time (GTT) variables are calculated as follows for each mode:

$$\mathrm{GTT}_{car} = \mathrm{TT}_{car,free} + \alpha_{congested} \times \mathrm{TT}_{car,congested} + \mathrm{TC}_{car}/VOT$$
$$\mathrm{GTT}_{PT} = \mathrm{TT}_{inv} + \alpha_{transfers} \times \mathrm{N}_{transfers} + \alpha_{wait} \times \mathrm{WaitT} + \alpha_{walk} \times \mathrm{WalkT} + \mathrm{TC}_{PT}/VOT$$
$$\mathrm{GTT}_{cycle} = \mathrm{TT}_{cycle,free} + \alpha_{congested} \times \mathrm{TT}_{cycle,congested}$$
$$\mathrm{GTT}_{walk} = \mathrm{TT}_{walk}$$

Table 3
Variables and constants descriptions.

Variables	Description	Constants	Value
$TT_{car,free}$	Car travel time under free flow conditions	$\alpha_{congested}$	1.5
$TT_{car,congested}$	Car travel time under congested conditions	VOT	92 DKK/h
TC_{car}	Car travel cost (car distance times 1.477DKK/km)	$\alpha_{transfers}$	9
TT_{inv}	Public transport in-vehicle time	α_{wait}	1.5
$N_{transfers}$	Public transport number of transfers	α_{walk}	1.5
WaitT	Public transport transfer waiting time		
WalkT	Public transport transfer walking time		
TC_{PT}	Public transport travel cost		
$TT_{cycle,free}$	Cycling travel time under free flow conditions		
$TT_{cycle,congested}$	Cycling travel time under congested conditions		
TT_{walk}	Walking travel time		

Table 4
Model estimates, with t-statistics given in brackets. All parameters except the starred one are significant at the 0.01 level for the MNL and the SBCM.

	MNL	BCM	$SBCM_{\lambda\to\infty}$	SBCM
Cost parameters ($\boldsymbol{\alpha}$)				
Alternative specific constants				
Car	–	–	–	–
Public transport	1.279 (17.31)	0.8363	0.6052 (9.175)	0.6083 (9.192)
Cycling	0.318 (5.681)	0.0162	−0.1512 (−4.450)	−0.1386 (−4.284)
Walk	1.052 (3.671)	−0.7499	−0.8133 (−8.993)	−0.8034 (−4.766)
Generalised travel time				
Car	0.0899 (33.19)	0.0619	0.0574 (21.59)	0.0566 (21.05)
Public transport	0.0197 (17.06)	0.0162	0.0174 (19.36)	0.0171 (18.33)
Cycling	0.1009 (43.90)	0.0852	0.0864 (41.78)	0.0851 (39.71)
Walk	0.0916 (14.04)	0.1034	0.1010 (31.66)	0.1020 (22.85)
Public transport variables				
Access time	0.0972 (15.04)	0.0781	0.0790 (14.93)	0.0790 (14.64)
Egress time	0.0828 (16.17)	0.0652	0.0667 (15.66)	0.0661 (15–24)
Scale (θ)	−1[a]	−1[a]	−1[a]	−1[a]
Relative bound parameter (φ)	–	4.369	5.812 (39.62)	6.675 (11.71)
Bound smoothing parameter (δ)	–	–	0.3714 (2.780)	0.1314 (1.953)
Maximum utility smoothing parameter (λ)	–	–	–	4.0274 (5.699)
Final LL	**−11,096.8**	**−10,782.7**	**−10,764.9**	**−10,754.7**
BIC	**22,232.9**	**21,609.3**	**21,578.8**	**21,562.7**
Adj. ρ^2	**0.579**	**0.591**	**0.592**	**0.592**
Number of parameters	**9**	**10**	**11**	**12**
Alternatives cut by bound	**0%**	**28.7%**	**22.8%**	**19.4%**

[a] The scale parameter is fixed to −1.

Table 3 gives the variable and fixed coefficient description and values. The values for $\alpha_{transfers}$ is extracted from Nielsen et al. (2021), while the values for $\alpha_{congested}, \alpha_{wait}, \alpha_{walk}$ are extracted from Hallberg et al. (2021). The car travel cost per kilometre and the Value of Time are extracted from the Danish transport ministry,[2] which were also used in Başaran et al. (2025).

Additionally, availability constraints have been added for car and bicycle trips, for which the respondent must possess a car with a driving license and a bicycle, respectively.

7.1.2. Results

The estimation results are given in Table 4, where we estimate MNL, the BCM, the SBCM with a free δ parameter and fixed λ parameter set to infinity ($SBCM_{\lambda\to\infty}$), and the SBCM with free δ and λ parameters. Below, we analyse different aspects of the estimation results. As the utility functions are not normalised, the scale parameter θ is not estimated (i.e., fixed to −1) for every model.

Model fit: Since the different models have different numbers of parameters to estimate, model fit is assessed according to the Bayesian Information Criterion (BIC) penalised likelihood criteria and Adjusted ρ^2. As can be seen, the BCM provides a considerably better fit to the data than MNL due to the cutting-off of mode alternatives (see below). Interestingly, the SBCM provides a marginally better fit to the data than the BCM, even when considering the number of parameters, with SBCM with free δ and λ parameters providing the best fit. We attribute this to the smoothness of the SBCM around the bound and reference utility.

[2] TERESA; https://www.cta.man.dtu.dk/modelbibliotek/teresa.

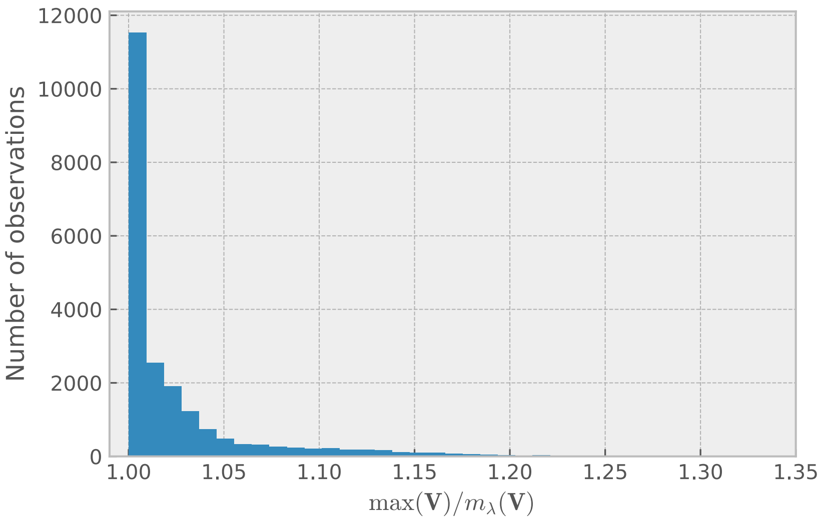


Fig. 9. Ratio of the maximum deterministic utility and the smoothened one.

Table 5
Percentage of available alternatives cut out by each model bound.

	Car trips	Public transport trips	Cycling trips	Walking trips
BCM	0%	16.8%	8.67%	77.2%
$\text{SBCM}_{\lambda\to\infty}$	0%	7.44%	2.08%	71.1%
SBCM	0%	3.17%	0.306%	64.7%

Model behavioural interpretation: All the model estimates make intuitive sense regarding the signs (see, e.g., Prato et al. (2017), for an analogous case study in the Copenhagen area). It is interesting to note that, according to every model, the public transport generalised travel time parameter is much less negative than the other modes, which may be linked to the inclusion of the access and egress times, as they are evaluated much more negatively than the GTT. Public Transport has the most negative ASC for every model, implying an inherent disutility of taking public transport.

The parameter λ is estimated to be around 4.02, implying that the reference alternative deterministic utility is slightly lower than the maximum deterministic utility in the choice set. Fig. 9 plots the histogram of the ratio $\max(\mathbf{V})/m_\lambda(\mathbf{V})$ over the observed choices. As can be seen, most of them are close to 1 (meaning there is almost no approximation error). There are a few observation with slight deviations, which mostly happens for short trips.

Interestingly, the SBCM with $\lambda \to \infty$ estimates a larger bound parameter than the BCM. This is likely due to the low δ parameter that creates the smooth shape, as highlighted in Fig. 7. Furthermore, the SBCM with a free λ parameter estimates a larger bound parameter than with $\lambda \to \infty$, which is likely due to the much lower bound smoothing parameter (δ) estimated for this model, meaning an even smoother model around the bounds. The consequence is that, across all choice situations, fewer alternatives are cut by the bound for SBCM than the BCM, i.e. 19.4 to 22.8% rather than 28.7%. Table 5 displays for each model the proportion of alternatives cut out by the bounds for each mode. As shown, the cut-offs are mainly composed of walking trips, which are often too long to be considered an alternative by travellers. Notably, both models cut off fewer cycling trips than public transport trips, implying that a larger proportion of trips in the Copenhagen area are considered feasible by bicycle than by public transport.

Table 6 shows the average attributes of the alternatives for each of the modes when that mode is cut off by the SBCM bound. These results suggest that the walk and bicycle alternatives are mainly cut off because the walking and cycling travel times are very long, rather than the car and public transport times being quick. We also observed that when the cycling alternative was cut off, so was the walking alternative, which makes sense as cycling is always quicker than walking, and the routes are similar. Interestingly, there were no cases where the cycling and public transport alternatives were cut off, implying that the bad performance of cycling and public transport are not correlated. As shown in Table 6, cut-off cycling alternatives are mainly overlong, while cut-off public transport alternatives are short trips in poorly connected areas. In these cases, it can, in generalised travel time, be five times faster to cycle, ten times faster to drive, and also faster to walk.

Estimates efficiency: The SBCM's differentiability allows us to calculate the t-statistics of estimated parameters, which evaluate estimate efficiency. These t-statistics are provided in Table 4 in brackets next to each parameter, calculated analytically using the Hessian matrix from Appendix C. As can be seen, the t-statistics are all over 1.96, meaning that every estimated parameter is statistically significant. It is a great advantage of the SBCM over the BCM that one obtains such information.

Table 6
Average attributes when an alternative is cut off by the SBCM bound.

Cut-off mode	GTT_{car}	GTT_{pub}	Acc	Egr	GTT_{cycle}	TT_{walk}
Public transport	7.46	60.49	9.74	12.11	16.33	50.66
Bicycle	32.48	146.2	7.92	9.72	143.1	401.4
Walk	20.87	73.07	7.83	9.45	53.00	156.57

Table 7
Aggregate point elasticities output by the MNL. The bold cells present direct elasticities, and the other ones are cross-elasticities.

Mode	GTT_{car}	GTT_{pub}	Acc	Egr	GTT_{cycle}	GTT_{walk}
Car	**−0.323**	0.140	0.065	0.063	0.176	0.011
Public transport	0.628	**−0.588**	**−0.311**	**−0.297**	0.521	0.045
Cycle	0.269	0.206	0.132	0.122	**−1.086**	0.060
Walk	0.160	0.224	0.194	0.182	0.558	**−3.490**

Table 8
Aggregate point elasticities output by the SBCM. The bold cells present direct elasticities, and the other ones are cross-elasticities.

Mode	GTT_{car}	GTT_{pub}	Acc	Egr	GTT_{cycle}	GTT_{walk}
Car	**−0.413**	0.137	0.061	0.058	0.250	0.0031
Public transport	0.633	**−0.579**	**−0.292**	**−0.273**	0.497	0.070
Cycle	0.554	0.206	0.126	0.114	**−1.375**	0.115
Walk	0.201	0.222	0.180	0.164	1.668	**−4.531**

Aggregate elasticities: The differentiability of the SBCM allows us to calculate aggregate elasticities. For instance, we can calculate how much, on average, an increase in Public Transport in-vehicle time will affect the choice probabilities of all the transport modes (and thus the predicted modal share). The aggregate elasticities use the point elasticities formulas from Eqs. (24), (25) and (27) for the SBCM. Elasticities from the MNL model are given in Table 7, and the SBCM elasticities are given in Table 8.

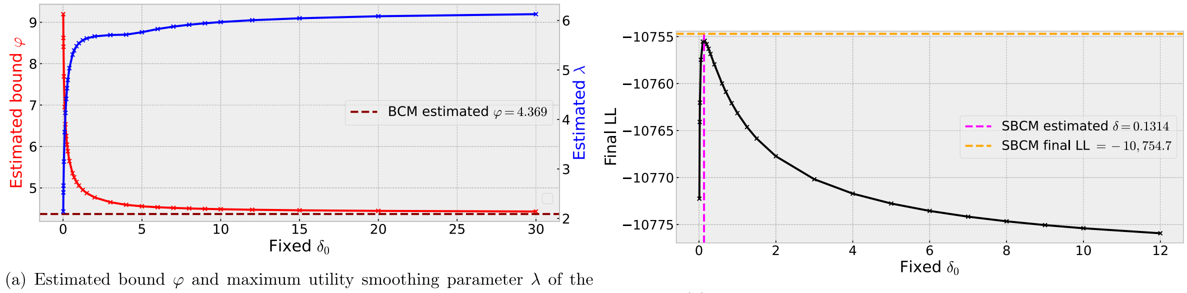
The two models output similar aggregate point elasticities, which make sense in size and magnitude. We can see, for instance, that relative changes in Generalised Travel Time (GTT) affect the choice probabilities of slow modes (cycling and walking) particularly. For instance, the SBCM outputs that a minor relative increase by a factor $\Delta > 1$ of the bicycle GTT will decrease the bicycle market share by 1.375Δ %, while increasing the modal share of all the other modes ($+1.668\Delta$ % for Walking, $+0.497\Delta$ % for Public Transport, and $+0.250\Delta$% for Car).

Some interesting differences can be found between MNL and SBCM elasticities. For instance, the elasticity of car probabilities to walking travel time is around four times smaller for the SBCM than MNL. This is because for most choice situations (11,700 out of 12,363 car choices, i.e. 94.6%), when the car is the chosen mode, walking was excluded from the consideration set. Hence, a marginal increase or decrease in walk travel time has zero impact on the car choice probabilities. This difference suggests that MNL overestimates the impact of walking travel time on car choice probabilities. We also observe that a marginal increase in cycling travel time has a much greater impact on the walking predicted share according to the SBCM (elasticity of 1.668) than the MNL (elasticity of 0.558), suggesting a large substitution of cycling trips to walking trips. Further analysis showed that observations that led to this massive increase in elasticities are the ones for which the car was not an available mode and for which public transport was not an attractive option compared to cycling. This suggests that a change in cycling travel time has a much larger impact on the walk modal share in these situations than MNL outputs and, thus, a larger predicted impact on behavioural change in the case of a pro-cycling policy.

7.1.3 Influence of the smoothing parameters

Here we analyse the estimation of the smoothing parameters of the SBCM, i.e. the bound smoothing parameter δ, and the maximum utility smoothing parameter λ. To do this, we estimate the SBCM for different fixed values of δ, and then for different fixed values of λ.

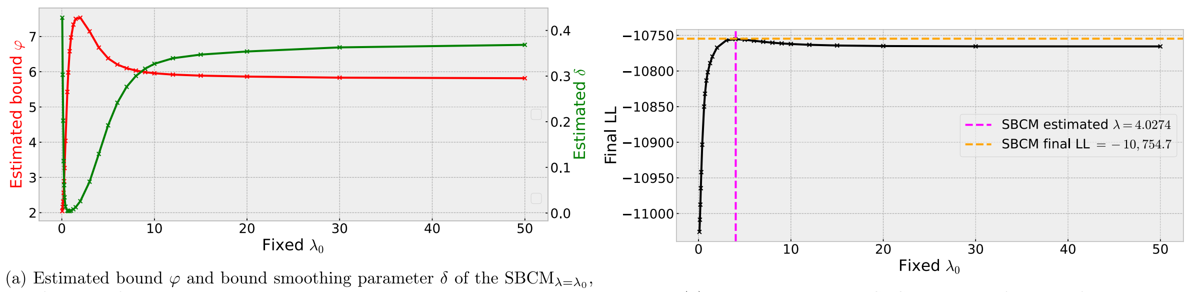
Influence of the bound smoothing parameter δ: We observed in the case study (Table 4) that including the bound smoothing parameter increased the bound parameter estimate φ. This intuitively makes sense, as smoothing the choice probabilities around this bound also decreases the likelihood of choices whose relative utility is close to this bound. To test for the influence of δ on the estimated bound φ and the smooth maximum parameter λ, we estimated several $SBCM_{\delta=\delta_0}$ with different fixed values for the smoothness parameter $\delta = \delta_0$ from 0.01 to 30, and observed how the bound parameter estimate, maximum utility smoothing parameter estimate, and model fit evolved. Fig. 10(a) plots how the relative bound parameter estimate φ and maximum utility smoothing parameter estimate λ vary for different fixed settings of the bound smoothing parameter δ. Fig. 10(b) plots the log-likelihood value. We observe that the estimated bound φ decreases with increasing δ_0, and collapses to the BCM relative bound. This was expected as the BCM is equivalent to $SBCM_{\delta\to\infty}$. We similarly see an increase in model fit until reaching the estimated $\delta = 0.371$ from Table 4, and then a decrease that tends to the BCM final likelihood. We also observed that increasing the bound smoothing parameter increased the estimated maximum utility smoothing parameter λ.



(a) Estimated bound φ and maximum utility smoothing parameter λ of the $\mathrm{SBCM}_{\delta=\delta_0}$, as a function of δ_0

(b) Final Log-Likelihood $\mathrm{SBCM}_{\delta=\delta_0}$, as a function of δ_0

Fig. 10. Influence of a fixed bound smoothing parameter δ on the other estimated parameters.



(a) Estimated bound φ and bound smoothing parameter δ of the $\mathrm{SBCM}_{\lambda=\lambda_0}$, as a function of λ_0

(b) Final Log-Likelihood $\mathrm{SBCM}_{\lambda=\lambda_0}$, as a function of λ_0

Fig. 11. Influence of a fixed maximum utility smoothing parameter λ on the other estimated parameters.

Influence of the maximum utility smoothing parameter λ: The maximum utility smoothing parameter λ influences the goodness of the approximation of the max operator by the Boltzmann operator m_λ. At the SBCM estimated parameter value ($\lambda = 4.0274$), the goodness of that approximation varies between observations. In general, observations with larger magnitudes of utility (i.e., long trips) get better approximations than short trips, and the relative error varies between $10^{-12}\%$ and 10%. In general, as the Boltzmann operator underestimates the maximum, we would expect that a lower value of λ would also decrease the bound, as the reference alternative utilities are underestimated. To test for the influence of δ on the estimated bound parameter φ and the maximum utility smoothing parameter λ, we estimated $\mathrm{SBCM}_{\lambda=\lambda_0}$ several times with different fixed values for the smoothness parameter $\lambda = \lambda_0$ from 0.01 to 50 and observed how the estimated bound and model fit evolved. Fig. 11(a) plots how the relative bound parameter estimate φ and bound smoothing parameter estimate δ vary for different fixed settings of the relative utility smoothing parameter λ. Fig. 11(b) plots the log-likelihood value. We observe that the interaction between λ and the estimated δ, and hence with the estimated bound, is not monotonous. We see an increase of the bound and a decrease of δ up to a critical value around $\lambda_0 = 2$, and a reverse trend for larger λ_0 values. This interaction is thus complex and hard to interpret.

7.2 Bicycle route choice in the Greater Copenhagen area

The second case study models cyclists' route choices in the Copenhagen Metropolitan area.

7.2.1 The data

The case utilised a large-scale crowd-sourced data set of bicycle GPS trajectories received from Hövding. The original dataset covers the entire Greater Copenhagen Area (see Fig. 12) in the period from the 16th September 2019 until 31st May 2021. For a detailed description of the data, the bicycle network, and the algorithms applied for data processing, we refer to Łukawska et al. (2023). The final dataset for model estimation consists of a subset of this dataset containing 4134 trips made by 4134 cyclists.

The cyclable network can be modelled as a directed graph $G = (B, A)$ where A is the set of links and B is the set of nodes. The network size is large, with $|B| = 420{,}973$ and $|A| = 324{,}492$. The network data was collected from Open Street Map (OSM[3]). The attributes of link $a \in A$ are as follows:

- L_a (km): Link length
- E_a (m): Link elevation gain when steepness > 3.5%
- No_a (km): Link length without bike infrastructure

[3] www.openstreetmap.org.

Fig. 12. Heatmaps of anonymised GPS trajectories from Hövding.

- S_a (km): Link length on a non-asphalt surface (i.e. gravel, cobblestones)
- W_a (km): Link length on wrong ways (cycling against traffic).

These attributes are stored in a travel utility attribute vector $\mathbf{T}_a = (L_a, E_a, \mathrm{No}_a, S_a, W_a)$. For a route i using a set of links $A_i \subseteq A$, these attributes are link-additive, so that the vector of travel cost attributes of route i is defined as $\mathbf{x}_i = \sum_{a \in A_i} \mathbf{T}_a$.

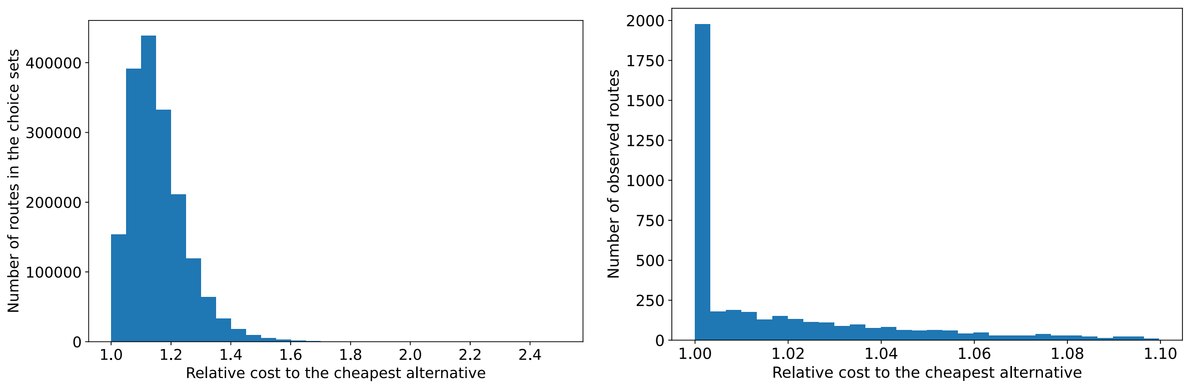
7.2.2 Choice set generation

Due to the extensive case size, it was not feasible to enumerate the universal set of routes between Origin and Destination (OD). Thus, as discussed in the introduction, we generated a representative universal choice set using practically motivated criteria. This choice set generation algorithm used a stochastic simulation approach (Nielsen, 2004; Bovy and Fiorenzo-Catalano, 2007), drawing a large number (10,000) of routes from the network between each OD based on randomly simulated link lengths. These lengths were Normally distributed around the actual lengths, i.e., for each $a \in E$, we drew a new value $\hat{L}_a \sim \mathcal{N}(L_a, \sigma L_a)$, where σ is a dispersion parameter around the initial length. In this study, we used a value of $\sigma = 0.5$. These routes were then filtered using a local optimality criterion (Abraham et al., 2013; Fischer, 2020), defined as the minimum length of a subpath that is not the shortest path. This criterion constrains the presence of small detours on routes and their mutual overlap. Fig. 13 shows the distribution of the relative deviation in utility (utility/max(utility)) of the generated and observed routes. As can be seen, a large proportion of the observations took the best route, and the maximum relative deviation in utility from the observations was around −1.1, meaning that the worst observation had a utility 1.1 times worse than the best alternative. In contrast, many routes were generated with a relative deviation in utility smaller than −1.1, suggesting that many of the generated routes in the representative universal choice sets may be cut off.

7.2.3 Results

Here we present the results from estimating the MNL, BCM, SBCM, BPS, and SBPS models, each with different parameters. For all models, there is a set of attribute parameters to estimate: $\alpha = (-1, \alpha_E, \alpha_{\mathrm{No}}, \alpha_S, \alpha_W)$, such that the linear utility function for a route $i \in C$ is $V_i = \boldsymbol{\alpha}^\top \mathbf{x}_i$, where $\mathbf{x}_i$ is the previously defined vector of attributes of route i defined in Section 7.2.1. Note that the utilities are normalised to the Length attribute, meaning that the other parameters can be seen as marginal rates of substitution to Length (usually referred to a Value-of-distance space, see, e.g., Łukawska et al. (2024)). The other parameters are as follows:

- For all models, there is a scale parameter θ, which refers to the scale parameter of the Logistic (MNL), truncated logistic (BCM, BPS), and smooth truncated logistic (SBCM, SBPS) distributions.
- For the BCM, SBCM, BPS, and SBPS models, there is a relative utility bound parameter φ, which is linked to the truncation/cutoff value Ψ_C of each choice set C ($\Psi_C = \varphi \max_{j \in C}(V_j)$ for the BCM, BPS, and $\Psi_C = \varphi m_\lambda(\mathbf{V})$ for the SBCM and SBPS).
- For SBCM and SBPS, there is a bound smoothing parameter δ to shape the smooth truncated logistic distribution (the larger δ, the smoother the distribution).



(a) Generated routes relative cost distribution (b) Observed routes relative cost distribution

Fig. 13. Relative costs distribution of the generated choice set and observed routes, using the MNL estimates from Table 9.

Table 9
Uncorrected model estimates. The t-statistic (i.e. the coefficient divided by its standard error) is given in brackets for each doubly-differentiable model. For these models, all the parameters are significant at the 0.01 level.

Model	MNL	BCM	SBCM$_{\lambda\to\infty}$	SBCM
Utility parameters (*α*)				
Length	−1[a]	−1[a]	−1[a]	−1[a]
Elevation gain	−0.0037 (2.031)	0.0846	−0.0044 (3.894)	−0.0044 (3.863)
No Bike infrastructure	−0.1808 (16.11)	4.302	−0.1652 (18.38)	−0.1678 (18.45)
Non-smooth surface	−0.1936 (43.81)	4.326	−0.1857 (49.72)	−0.1856 (49.64)
Wrong way	−0.3319 (40.15)	7.875	−0.3386 (45.77)	−0.3407 (45.93)
Scale (θ)	28.54 (64.40)	25.46	23.17 (52.47)	22.81 (51.63)
Relative bound parameter (φ)	–	1.110	1.1535 (208.1)	1.1522 (213.3)
Bound smoothing parameter (δ)	–	–	0.1496 (5.078)	0.1347 (5.230)
Maximum utility smoothing parameter (λ)	–	–	–	4.586 (6.711)
Final LL	**−11,075.6**	**−10,777.8**	**−10,705.3**	**−10,702.6**
BIC	**22,169.3**	**21,577.7**	**21,435.9**	**21,434.2**
Adj. ρ^2	**0.5135**	**0.5265**	**0.5296**	**0.5297**
N params	**5**	**6**	**7**	**8**
Routes cut by bound	**0%**	**66.6%**	**49.0%**	**48.8%**

[a] The parameter associated with Length is set to −1.

- For SBCM and SBPS, there is a maximum utility smoothing parameter λ controlling the quality of the smooth approximation of the max function for the reference utility.
- For BPS and SBPS, there is a Path-Size scaling parameter η that influences the weight of the Path-Size correction in penalising the utility function.

Tables 9 and 10 displays the model parameter estimates. For the smooth models, we present two versions: one for which λ is fixed to an arbitrarily large value (models referred to as SBCM$_{\lambda\to\infty}$ and SBPS$_{\lambda\to\infty}$), and one for which λ is freely estimated via MLE. Table 9 also displays t-statistics (where possible) and the percentage of routes cut by the bound.

All estimated parameters are significant, with the bound being, by far, the most significant parameter. Every bounded model allocates zero probabilities to between 49% and 67% of the choice set, which means that according to these models, around half of the generated routes in the choice sets are not even considered by cyclists due to their too-high generalised cost.[4] This is most likely the main reason bounded models perform better than the MNL, as the MNL must allocate a non-zero probability to all those routes. The SBCM also outperforms the standard BCM in terms of fit. The SBCM relative cost bound is estimated higher, probably because of the smoothness of the probability function (there is no fast increase of the choice probabilities around the bound). Still, this smoothness seems to represent choice behaviour better. Finally, accounting for the inherent correlation between routes with the BPS and SBPS models leads to large improvements in model fit. This justifies why accounting for the correlation between routes is crucial when modelling their choice probabilities. The estimation of λ, while leading to significant parameters and a small log-likelihood improvement, does not provide further behavioural insights. The goodness of the approximation of the max function

[4] This percentage is highly dependant on the choice set generation method. For instance, methods that generate routes with a higher variance or number of draws (for stochastic methods) or with different deterministic criteria may be more prone to contain many unrealistic alternatives that the bound would cut out.

Table 10
Path-Size corrected model estimates. The t-statistic (i.e. the coefficient divided by its standard error) is given in brackets for each doubly-differentiable model. For these models, all the parameters are significant at the 0.01 level.

Model	BPS	$\mathrm{SBPS}_{\lambda\to\infty}$	SBPS
Cost parameters ($\boldsymbol{\alpha}$)			
Length	−1*	−1*	−1*
Elevation gain	−0.0036	−0.0038 (4.172)	−0.0038 (1.990)
No Bike infrastructure	−0.1592	−0.1476 (18.99)	−0.1491 (18.86)
Non-smooth surface	−0.1531	−0.1503 (48.43)	−0.1498 (47.30)
Wrong way	−0.2670	−0.2617 (44.12)	−0.2623 (43.21)
Scale (θ)	14.71	14.495 (53.15)	14.396 (50.71)
Path-Size coefficient (η)	1.643	1.632 (42.41)	1.629 (42.13)
Bound (φ)	1.105	1.123 (249.7)	1.123 (224.2)
Bound smoothing parameter (δ)	–	1.093 (3.612)	0.894 (3.173)
Maximum utility smoothing parameter (λ)	–	–	8.110 (3.814)
Final LL	**−9910.1**	**−9882.5**	**−9880.4**
BIC	**19,845.2**	**19,793.9**	**19,793.4**
Adj. ρ^2	**0.5646**	**0.5657**	**0.5658**
N params	**7**	**8**	**9**
Routes cut by bound	**65.8%**	**55.7%**	**54.7%**

varies: for short trips (less than 1 km), there is a significant difference between the $\max(\mathbf{V})$ and $m_\lambda(\mathbf{V})$ (up to 15%), while for longer trips, the difference is negligible (down to 10^{-15}%).

Figs. 14(a) and 14(b) show how the choice probabilities look in a binary case of two alternatives 1 and 2 with respective deterministic utilities V_1, V_2 with the model estimates from Table 9. We plotted the MNL, BCM, and SBCM choice probabilities as a function of the cost of Alternative 2, given that Alternative 1's cost is fixed.

In Fig. 14(a), $V_1 = 1$. This plot shows that there is a slight asymmetry on how the BCM and SBCM choice probabilities evolve with V_2 (in the sense that $\mathbb{P}(2|V_2 = V_1 - \Delta) \neq \mathbb{P}(1|V_2 = V_1 + \Delta)$ for $0 < \Delta < (\varphi - 1)V_1$). Moreover, the low smoothness parameter $\delta = 0.149$ makes the choice probabilities get faster to zero for the SBCM than for the BCM. While the SBCM has a lower scale parameter than the MNL and BCM, its choice probabilities have the steepest slope for V_2, because of the distributional assumptions of the model. In Fig. 14(b), $V_1 = 2$. In this case, the relative utility bound φ is located much further away from V_1, so the choice probabilities have a sigmoid-resembling shape (with different slopes given by the different estimated scales for each model).

In Fig. 15, we plot the distribution of the error terms difference ε_i for the different estimated models in Table 9, assuming a reference utility being equal to −1. This plot shows the difference between the distributions for short trips. The bound is slightly larger for the SBCM than the BCM, which makes sense with the previous case study experiments.

7.3 Public transport route choice in Copenhagen

In the third case study, we estimated the MNL, BCM, and SBCM as route choice models, but in this case, they were based on the Greater Copenhagen Region's large-scale multimodal public transport network.

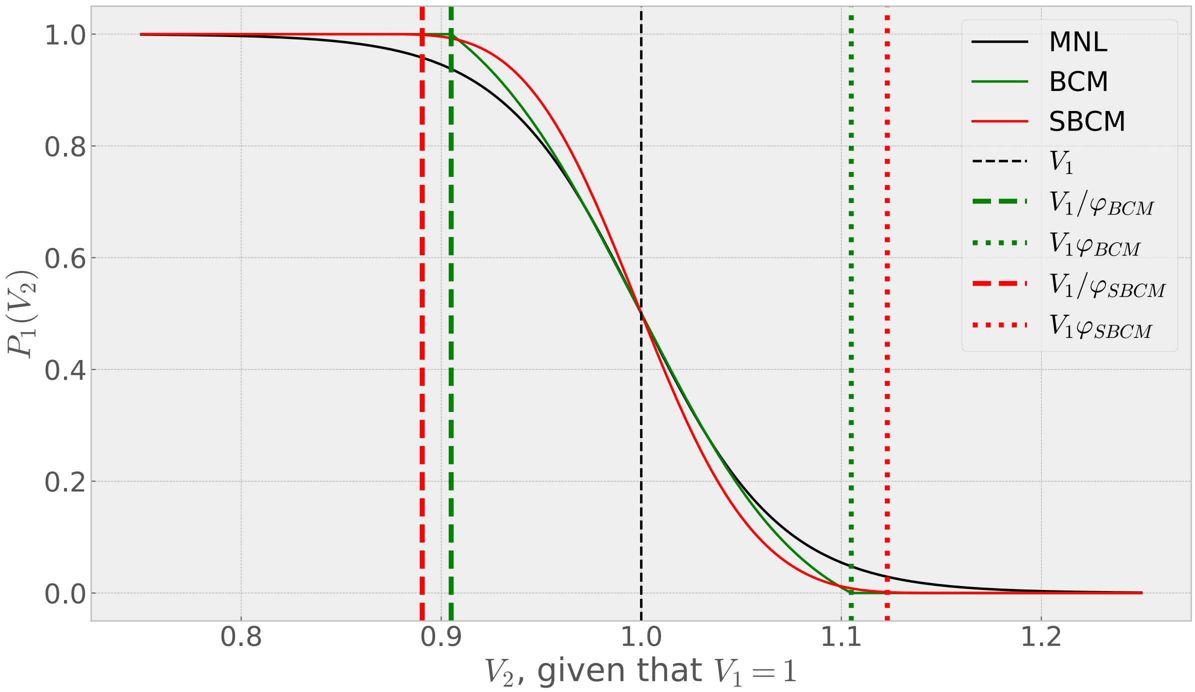
7.3.1 Data

We use a multimodal public transport dataset. A thorough presentation of this dataset can be found in Nielsen et al. (2021). Anderson (2013) collected the 4810 observed routes as part of the Danish National Travel Survey. These observations are separated into two subsets: work-related trips (2553 observations) and leisure trips (2257 observations), and separate models were estimated for these two datasets. The representative universal choice set was generated using a Doubly-Stochastic method (Nielsen, 2004). The dataset contains the travel time in each transport mode: bus, commuter train (S-train), Metro, Regional and intercity train, and local train. It also contains transfer components (number of transfers, waiting time and walking time), access and egress time, and highest headway in the trip. The model does not include the Path-Size correction term, as Nielsen et al. (2021) did not find it a significant explanatory variable.

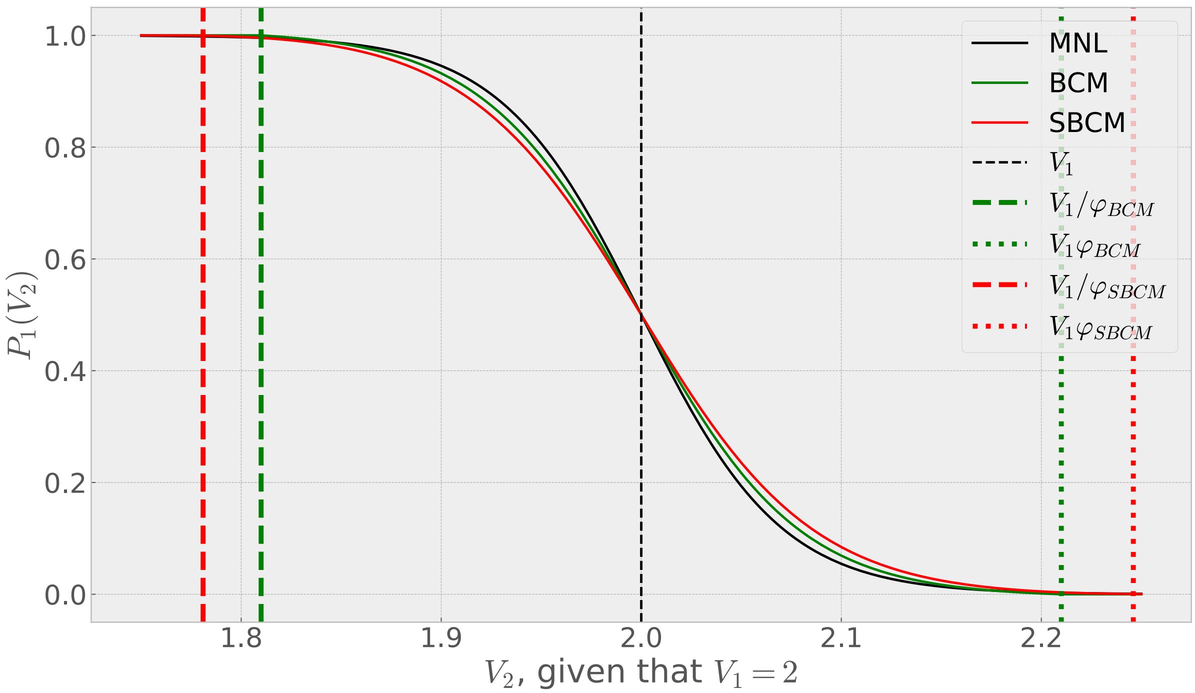
7.3.2 Results

We estimated a MNL, a BCM and a SBCM for both trip purposes. Preliminary results showed that the reference alternative smoothness parameter λ was estimated to $+\infty$. Thus, we fixed it to a large number ($\lambda = 100$) and estimated the model $\mathrm{SBCM}_{\lambda=100}$ where λ is a hyperparameter. Similarly to the bicycle route choice case study, the utilities have been scaled to Bus In-vehicle time, which implies that the other cost parameters translate the relative sensitivity to attributes to Bus In-Vehicle time. Table 11 gives the model estimation results. The t-statistics for each smooth model (the MNL and the SBCM) were calculated analytically using the likelihood Hessian matrix. The SBCM, for instance, outputs that the decision-makers are, on average, willing to trade 1 min of S-train for 0.7385 min of Bus without a change in utility.

The model estimates make sense in magnitude and sign, corroborating with Nielsen et al. (2021). Similarly to the previous case studies, the BCM relative utility bound improves the model fit significantly. However, in this case study, neither of the two



(a)



(b)

Fig. 14. (a) Binary choice probability of alternative 1 according to the MNL, BCM and SBCM estimates parameters, with the utility of alternative 1, V_1 fixed to 1. (b) As for (a) but with V_1 fixed to 2.

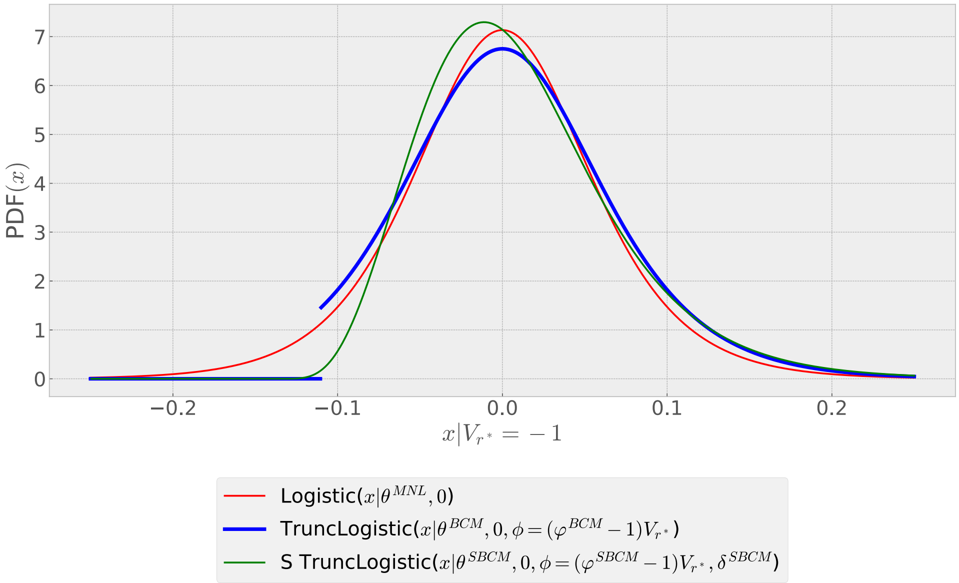


Fig. 15. Distribution plots for the model estimated parameters, assuming that the reference deterministic utility is -1.

Table 11
Model results for Work and Leisure trips, the t-statistics are given in brackets. The smoothness parameters δ are both insignificant at the .05 level. All the other parameters are significant at the .001 level.

Trip purpose	Work			Leisure		
Model	MNL	BCM	SBCM$_{\lambda=100}$	MNL	BCM	SBCM$_{\lambda=100}$
Cost parameters (θ)						
In-vehicle time						
Bus	−1*	−1*	−1*	−1*	−1*	−1*
Metro	−0.3897 (6.81)	−0.3820	−0.3826 (6.96)	−0.3834 (5.00)	−0.3322	−0.3481 (5.72)
Reg. and intercity train	−0.8868 (22.10)	−0.8848	−0.8813 (21.65)	−0.8826 (17.52)	−0.9641	−0.9660 (17.44)
S-Train	−0.7478 (30.70)	−0.7414	−0.7385 (30.36)	−0.7393 (24.86)	−0.7173	−0.7175 (27.05)
Local train	−0.9056 (12.60)	−0.9104	−0.9130 (12.09)	−0.9116 (6.38)	−0.7925	−0.7944 (5.65)
Transfer components						
Nb of transfers	−8.2194 (25.38)	−8.0394	−7.9844 (24.24)	−7.9935 (23.17)	−8.3266	−8.2231 (23.65)
Transfer walk time	−0.6767 (8.33)	−0.6606	−0.6606 (7.18)	−0.6607 (7.66)	−0.7216	−0.7176 (8.03)
Transfer wait time	−0.1542 (5.16)	−0.1555	−0.1571 (5.13)	−0.1554 (4.64)	−0.1494	−0.1520 (4.87)
Other components						
Access time	−1.6567 (26.60)	−1.6362	−1.6333 (26.16)	−1.6312 (19.07)	−1.7260	−1.7235 (19.20)
Egress time	−1.4705 (19.89)	−1.4556	−1.4540 (20.45)	−1.4548 (17.79)	−1.4694	−1.4593 (19.34)
Trip highest headway	−0.4716 (9.66)	−0.4813	−0.4825 (10.28)	−0.4822 (9.16)	−0.4314	−0.4353 (3.79)
Scale (θ)	0.3534 (37.06)	0.3372	0.3272 (30.59)	0.3358 (33.77)	0.3045	0.2994 (29.37)
Bound (φ)	–	1.526	1.686 (17.89)	–	1.533	1.722 (28.06)
Smoothness parameter (δ)	–	–	*0.1738 (0.9592)*	–	–	*0.1730 (1.873)*
Final LL	**−2391**	**−2373**	**−2371**	**−2623**	**−2579**	**−2573**
BIC	**4868**	**4840**	**4844**	**−5331**	**5251**	**5239**
Adj. ρ^2	**0.804**	**0.805**	**0.806**	**0.745**	**0.749**	**0.751**
N params	**11**	**12**	**13**	**11**	**12**	**12**
Routes cut by bound	**0%**	**90.2%**	**86.7%**	**0%**	**88.9%**	**85.5%**

introduced smoothness parameters was found to be significant, even though the parameter δ slightly improves the model fit in the Leisure case. This implies that the shape of the BCM choice probabilities is more suited to explain the route choices made by this dataset's decision-makers. The presence of the two smoothness parameters is still important, as it allows the choice probabilities differentiability. However, as a model hyperparameter, the smoothness parameter δ could also be fixed to a large positive number. Interestingly, the bounds cut off between 85% and 90% of the generated routes, implying that a large part of the sampled routes from the universal choice set were unrealistic. The BCM and SBCM estimates are very similar, with a slightly higher estimated relative cost bound for the smooth version (as also found with the Hövding dataset). This difference may be attributed to the smoothness of the probability function, as choice probabilities increase less fast than for the BCM around the bound. As we saw for the Hövding dataset, the SBCM significantly improves the model fit to the data.

8 Discussion

8.1 Summary of contributions

In this paper, we have advanced the field of one-stage choice set formation models, which provide a behavioural and computational advantage over no-choice set formation and two-stage choice set formation models. For reasons discussed in the paper, the BCM (Watling et al., 2018) is a particularly attractive one-stage choice set formation model as it imposes hard, pervasive, compensatory, endogenous, and continuous cutoffs upon alternatives to determine implicit consideration choice sets consistent with the choice probabilities.

However, as demonstrated in the paper, the BCM is not differentiable. We have, therefore, developed the Smooth BCM (SBCM) to address this shortcoming. By relying on a new smooth bounded-support distribution for the random error terms and smoothly approximating the reference utility, the closed-form SBCM choice probabilities are infinitely differentiable (smooth). The core features of the BCM are preserved. For example, the MNL model can be approximated as the bound tends to infinity, and the bounding criterion is the same: alternatives receive zero probability if their utility violates the bound. The SBCM can also approximate the original BCM to arbitrary precision under a specific setting of the hyperparameters. The smoothness property of the SBCM guarantees the asymptotic normality of maximum likelihood parameter estimates and facilitates the calculation of likelihood gradients, the AVC matrix of the estimates, confidence intervals, and elasticities.

In choice contexts such as route choice, where there is considerable correlation between alternatives, this correlation should be accounted for Florian and Fox (1976). However, the BCM and SBCM do not account for correlation. Extending the BCM to account for such in the path choice context, Duncan et al. (2022) recently developed the Bounded Path Size (BPS) route choice model, which includes path size correction terms within the BCM probability relation to capture route correlation. In this paper, we have modified the BPS model analogously to how we modified the BCM to formulate the SBCM to formulate a smooth SBPS model.

Subsequently, we presented a MLE technique for the SBCM model. We proposed parameterisation techniques for the model parameters to avoid using box constraints on the parameter values and speed up the convergence of the estimation procedure. Since the SBCM log-likelihood function is not guaranteed to be concave, we explored the uniqueness of MLE solutions numerically, solving with different randomly generated initial conditions. No cases of multiple solutions were found.

To explore these features, we estimated the SBCM on three large datasets of observed choice behaviour from the Greater Copenhagen area: a mode choice dataset, a bicycle route choice dataset, and a public transport dataset. We calculated standard errors and marginal rates of substitution for the parameter estimates using the analytical gradient. Additionally, we calculated aggregate point elasticities for the mode choice case study and analysed the alternatives excluded by the model bound. Benchmarked against the MNL, the SBCM provided plausible results and considerably better fit the data in all the case studies. These fit improvements outweighed the additional parameters in all but one case, in which the BIC increased. We hypothesise that the smoother shape of the probability function is more suited to model choice behaviour. Upon the bicycle route choice application, we found that the SBPS model considerably outperformed both the MNL and the SBCM, highlighting the importance of accounting for correlation and unrealistic routes in generated choice sets.

An estimated bound implicitly identifies the alternatives individuals do not consider by giving them zero probability. This helps interpret an individual's consideration set. Interestingly, the number of excluded alternatives was high in the route choice case studies. This suggests that the stochastic choice set generation method generated many unrealistic alternatives to which the MNL cannot allocate zero probability. Failing to do so may have led to biased predictions, especially in large-scale case studies. Excluding unrealistic alternatives also affects the calculation of policy-related indicators, like elasticities, as these excluded alternatives do not influence their calculation. For instance, the SBCM suggested that the substitution between walking and using the car was overestimated by the MNL, as, in most cases, walking was not a realistic alternative when the car was chosen. Conversely, the substitution between walking and cycling was underestimated.

8.2 Conclusion

In conclusion, we believe the SBCM is a promising model. First, its closed-form probabilities, analytical gradients, and Hessian matrices make it easy to estimate, notably for large choice sets. This is not the case for every one-stage choice set formation model, which often requires advanced estimation techniques, such as solving fixed point problems or using Bayesian inference. In our case studies, the SBCM provides a much better fit to the data than the MNL model, likely because unrealistic alternatives are assigned zero probabilities and thus do not influence the model estimates. Secondly, the SBCM allows for a richer interpretation of the model estimates than the MNL for the analysis of zero-probability alternatives from the representative universal choice set and the BCM due to the possibility of analytical calculation of standard errors and elasticities. Furthermore, while the BCM has yet only been applied to car route choice, we have now shown its suitability to different choice situations.

8.3 Future research

There are several directions for future research. Firstly, it is possible to incorporate the SBCM smoothing techniques in the Conjunctive BCM (Rasmussen et al., 2024) and also in an analogous Disjunctive BCM that could be based on previous work on disjunctive models (Cazor et al., 2024). This would allow for comparing different decision rules in one-stage choice set formation. A method for assessing which choice set formation assumption from the literature is the most suitable depending on the case study could also be developed. This could be done either by comparing information criteria such as the BIC, or the model predictive ability on out-of-sample datasets.

Several other extensions of the SBCM could also enhance its versatility. For instance, a Nested SBCM for non-route choice contexts could be developed to relax the Independence of Irrelevant Alternatives (IIA) property and allow for correlation between alternatives. Another example could be to account for taste or bound heterogeneity by mixing some parameters with a discrete or continuous distribution.

Another stream of research could work further on the estimation procedure. A concern with the SBCM is its sensitivity to outliers, i.e., chosen alternatives that have a relative high utility compared to the best-performing alternative. During estimation, the bound is set to ensure that no observed alternative is entirely cut off (i.e., it always retains some likelihood). However, in out-of-sample validation, new outliers may emerge, potentially leading to poor overall model performance due to zero likelihood assignments. Alternative objective functions that are less sensitive to zero probabilities could be used for model assessment and possibly parameter estimation. Another approach involves systematically identifying outliers to better understand their impact. Cut-off issues may also arise due to insufficient heterogeneity in observed choice behaviour, highlighting the importance of ensuring that the estimation dataset is representative of the population. A possible solution could be to incorporate taste heterogeneity or bound heterogeneity, such as individual-specific bounds, or mixing parameters with a random distribution (McFadden and Train, 2000). This would allow for greater flexibility in capturing variation across decision-makers, potentially improving both estimation robustness and predictive performance.

CRediT authorship contribution statement

Laurent Cazor: Validation, Formal analysis, Writing – original draft, Software, Data curation, Visualization, Methodology, Conceptualization. **Lawrence Christopher Duncan:** Validation, Conceptualization, Supervision, Writing – review & editing, Methodology. **David Paul Watling:** Methodology, Writing – review & editing, Conceptualization, Supervision. **Otto Anker Nielsen:** Conceptualization, Writing – review & editing, Supervision. **Thomas Kjær Rasmussen:** Supervision, Conceptualization, Writing – review & editing, Project administration, Validation, Funding acquisition.

Declaration of competing interest

None

Acknowledgements

We thank the Independent Research Fund Denmark (DFF) for funding this research through the projects "Next-generation route choice models for behavioural realism and large-scale applications" (Grant ID: 0136-00242B), "Human energy expenditure in mobility" (Grant ID: 2067-00077B), and "Optimal tolls for reaching emissions goals: A novel behaviourally realistic and national-scale applicable road transport modelling system" (Grant ID: 0217-00173B). We also thank Gülin Goksu Basaran and the Easier project for her help with the Mode choice case study data set. We thank the Bikelonger project funder for the bicycle route choice case study dataset.

Appendix A. One-stage choice set formation models: a review

In this first Appendix, we explain and present one-stage choice set formation models. Under utility maximisation, if $C_n \subseteq C$ is the decision-maker n choice set, the choice probability of alternative i is given by:

$$P_n(i) = \Pr(U_{in} \geq U_{jn}, \forall j \in C_n) \tag{32}$$

where $U_{jn} = V_{jn} + \epsilon_{in}$ is the utility of alternative j for decision-maker n. Assuming the analyst knows C_n, it can be described by deterministic availability indicators (Bierlaire et al., 2010):

$$A_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is considered by individual } n \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

The choice model can be re-written:

$$P_n(i) = \Pr(U_{in} + \ln(A_{in}) \geq U_{jn} + \ln(A_{jn}), \forall j \in C) \tag{34}$$

To model for the analyst's lack of knowledge on the actual composition of C_n, Cascetta and Papola (2001) proposed to replace A_{in} by a penalty term $\phi_{in} \in [0, 1]$, which represents the probability that individual n considers alternative i. Several models from the literature explored different functional forms for the penalty term. This framework is analogous to the modelling framework presented by Brathwaite and Walker (2018), where the utility is transformed with a function S before calculating the choice probabilities using a logit formula. While several models below could be included within their framework, using $S(V_{in}) = V_{in} + \ln(\phi_{in})$, Brathwaite and Walker (2018) make the assumption that the function S of an alternative does not depend on the attributes of other alternatives (i.e., maintain the Independence of Irrelevant Alternatives, or IIA assumption). This is not the case of some of the presented models, such as the BCM (Watling et al., 2018), the SBCM (this paper), and the Conjunctive BCM (Rasmussen et al., 2024).

A.1. Swait (2001b) choice set formation model

First, Swait (2001b) model imposes cutoffs Ψ_k to the attributes k, meaning that if an alternative attribute x_{ink} fails to pass a cutoff value, its utility will be penalised. According to this model, the utility can be rewritten as a sum of subutilities for each attribute k, i.e., $V_{in} = \sum_k V_{ink}$, where V_{ink} is defined by:

$$V_{ink} = \theta_k x_{ink} + \gamma_k \max(0, \Psi_k - x_{ink})$$

which can also be expressed as:

$$V_{ink} = \begin{cases} \theta_k x_{ink} & \text{if } x_{ink} \geq \Psi_k \\ (\theta_k - \gamma_k) x_{ink} + \gamma_k \Psi_k & \text{otherwise} \end{cases} \tag{35}$$

$\gamma_k > 0$ is a penalty coefficient. $\gamma_k = +\infty$ represents a pure conjunctive behaviour (i.e., if an attribute does not meet the cutoff, the choice probabilities become zero). This utility can be rewritten $V_{ink} = \theta_k x_{ink} + \ln \phi_{ink}$, where $\phi_{ink} = \exp(\gamma_k \max(0, \Psi_k - x_{ink}))$. The total penalty for an alternative i is thus equal to $\phi_{in} = \prod_{k=1}^{K} \phi_{ink}$. A similar penalty can be given for upper attribute cutoffs.

A.2. The Constrained Multinomial Logit (CMNL) model

The CMNL (Martínez et al., 2009) is presented as a smooth version of Swait's model. It is noted that Paleti (2015) extended this model to better approximate Manski (1977)'s framework. The CMNL imposes upper and lower cutoffs on the attribute value (noted a_k and b_k), using a slightly different definition for ϕ_{in}: $\phi_{in} = \prod_{k=1}^{K} \phi_{ink}^{L} \phi_{ink}^{U}$, where:

$$\phi_{ink}^{U} = \frac{1}{1 + e^{\omega_k (x_{ink} - a_k + \rho_k)}}$$

$$\phi_{ink}^{L} = \frac{1}{1 + e^{\omega_k (b_k - x_{ink} + \rho_k)}}$$

ω_k is a scale parameter and ρ_k is a location parameter.

Attribute cutoffs are considered *soft*: as ϕ_{in} cannot be zero, a violation of the cutoff penalises the choice probability but cannot lead to a zero probability. This property may be problematic for some applications (e.g., route choice), where eliminating alternatives with certainty simplifies the choice set generation task. To our knowledge, two models have been developed that allocate zero probability to alternatives that do not meet a cutoff value.

A.3. Elrod et al. (2004) choice set formation model

Elrod et al. (2004) develops a "pervasive conjunctive + linear model" using a general nonrectangular hyperbola (GNH) value function. The word "pervasive" means that if an alternative is close to the cutoff value, it will also be penalised (in a continuous and differentiable way). According to the GNH model, the deterministic utility $V_{in} = \sum_k V_{ink}$, where V_{ink} is defined by:

$$V_{ink} = \begin{cases} \frac{-\gamma_k}{x_{ink} - \Psi_k} + \theta_k x_{ink} & \text{if } x_{ink} \geq \Psi_k \\ -\infty & \text{otherwise} \end{cases} \tag{36}$$

It follows that the GNH model penalty can be written as $\phi_{in} = \prod_{k=1}^{K} \phi_{ink}$, where $\phi_{ink} = \exp\left(\frac{-\gamma_k}{x_{ink} - \Psi_k}\right)$.

A.4. The Bounded Choice Model (BCM)

The BCM (Watling et al., 2018; Duncan et al., 2022) assumes that the (representative) universal choice set may contain unrealistic alternatives with large costs/utilities, following empirical observations from Watling et al. (2015), and that these never enter individuals' consideration sets. This model assumes that individuals consider an alternative if its utility or cost is within some bound of an imaginary reference alternative. It allocates zero probability alternatives by assuming that the difference of error terms between the utility of any alternative and one of the reference alternatives follows a truncated logistic distribution rather than a logistic distribution for the MNL. Its choice probabilities are defined as:

$$P_{in}^{\mathrm{BCM}} = \frac{\left(\exp(\theta(V_{in} - \Psi_n)) - 1\right)_+}{\sum_{j \in C} \left(\exp(\theta(V_{jn} - \Psi_n)) - 1\right)_+} \tag{37}$$

where Ψ_n is a cutoff value for the overall utility. Watling et al. (2018) defined this cutoff endogenously by assuming that it is related to the best-performing alternative in the choice set (i.e., $\Psi_n = \varphi \max_{j \in C} V_{jn}, \varphi > 1$ for a relative model, or $\Psi_n = \max_{j \in C} V_{jn} + \delta, \delta > 0$, for an absolute model). It is possible to show that the BCM is equivalent to a one-stage formation model whose modified deterministic utility $\tilde{V}_{in}$ is given by:

$$\tilde{V}_{in} = \begin{cases} \ln(1 - e^{\theta(V_{in} - \Psi_n)}) + \theta V_{in} & \text{if } V_{in} \geq \Psi_n \\ -\infty & \text{otherwise} \end{cases} \tag{38}$$

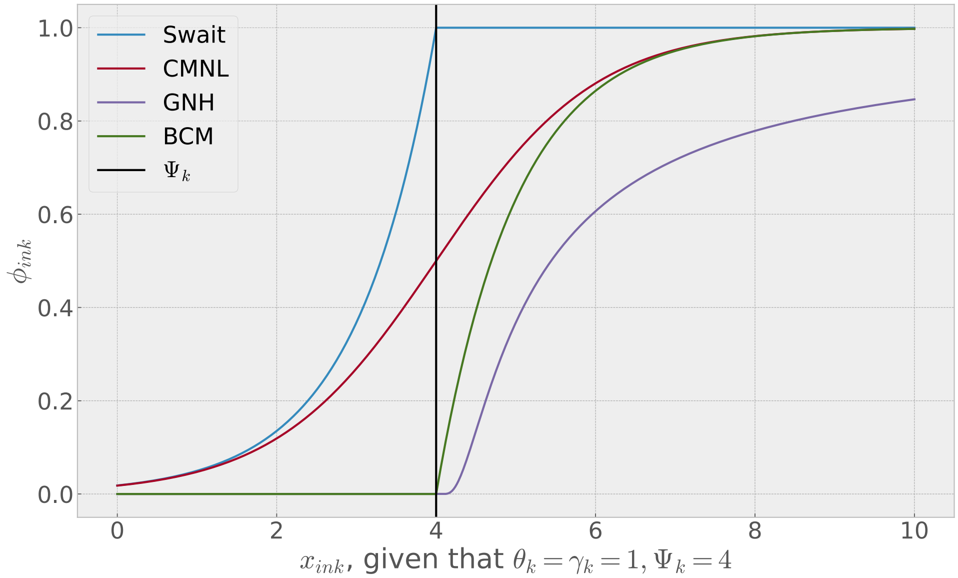


Fig. 16. Penalties for violating a lower bound for the different one-stage formation models.

The different penalties are plotted in Fig. 16. This figure showcases the different penalty properties. Firstly, we can see that Swait (2001a) and Martínez et al. (2009)'s CMNL penalties are "soft", i.e., that they never allocate zero probability to an alternative, regardless of how badly an attribute performs. Conversely, the GNH and BCM penalties can be considered "hard". Secondly, we observe that the CMNL, GNH and BCM penalties are "pervasive", meaning that alternatives whose attribute values are close to the cutoff are more penalised than those far. The above-mentioned properties are summarised in Table 1.

Appendix B. Proofs of the properties of the smooth truncated logistic distribution

Property 1. *The smooth truncated logistic distribution has bounded support on* $[-\phi, +\infty)$

Proof. For $x < -\phi, \exp(\theta(x+\phi)) - 1 < 0$, so $g_\delta(\exp(\theta(x+\phi)) - 1) = 0$. As $g_\delta(\exp(\theta\phi) - 1) > 0$, we have that $F_S(x|\theta, \phi, \delta) = 0$ □

Property 2. *The CDF of the smooth truncated logistic distribution is infinitely differentiable on* $\mathbb{R}$.

Proof. It is sufficient to prove that F_S is infinitely differentiable at the breakpoint $x = -\phi$, which is equivalent to proving the infinite differentiability of the function g_δ at the breakpoint $x = 0$. For $n > 1$, we note $g_\delta^{(n)}$ the nth order derivative of g_δ. We can prove by induction that there exists a polynomial P_n of degree $2n$, so that:

$$g_\delta^{(n)}(z) = \begin{cases} P_n\left(\frac{1}{z}\right)\exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

As, for any integer k, the function $z^k \exp(-z)$ tends to 0 when z tends to $+\infty$, it follows that, for any polynomial P, $P(z)\exp(-z)$ tends to 0 when x tends to $+\infty$, or equivalently, that $P\left(\frac{1}{z}\right)\exp\left(-\frac{1}{\delta z}\right)$ tends to 0 when z tends to 0. This ensures the continuity of the $g_\delta^{(n)}(z)$ at the breakpoint.

Consequently, F_S is infinity differentiable as a composition, sum, and ratio (which never equates zero) of infinitely differentiable functions. □

Property 3. *The smooth truncated logistic distribution collapses to the truncated logistic distribution as* $\delta = +\infty$.

Proof. First, we will prove that the g_δ functions uniformly converge to $(.)_+$ when $\delta \to +\infty$. We have that $\|(.)_+ - g_\delta\|_\infty = \max_{z>0}(z(1-\exp(-1/\delta z)))$, as the function is positive. The function $z \to z(1-\exp(-1/\delta z))$ is increasing, so $\max_{z>0}(z(1-\exp(-1/\delta z))) = \lim_{z\to+\infty} z(1-\exp(-1/\delta z))$. This limit can be calculated thanks to the l'Hopital's rule:

$$\lim_{z\to+\infty} z(1-\exp(-1/\delta z)) = \lim_{z\to+\infty} \frac{1-\exp(-1/\delta z)}{\frac{1}{z}} \tag{40}$$

$$=_{\text{L'Hopital}} \lim_{z\to+\infty} \frac{-\frac{\exp(-1/\delta z)}{\delta z^2}}{\frac{-1}{z^2}} \tag{41}$$

$$= \lim_{z\to+\infty} \frac{1}{\delta}\exp(-1/\delta z) \tag{42}$$

$$= \frac{1}{\delta} \tag{43}$$

Hence, $\|(.)_+ - g_\delta\|_\infty = 1/\delta \xrightarrow[\delta\to+\infty]{} 0$, meaning that g_δ uniformly converges to $(.)_+$ when $\delta \to +\infty$. This property implies that the C^∞ truncated Logistic distribution, when $\delta \to +\infty$, uniformly converges to the following distribution:

$$\lim_{\delta\to+\infty} F_S(x|\theta,\phi,\delta) = \frac{(\exp(\theta(x+\phi))-1)_+}{(\exp(\theta(x+\phi))-1)_+ + (\exp(\theta\phi)+1)_+} \tag{44}$$

For $x > -\phi$, we have that $(\exp(\theta(x+\phi))-1)_+ = \exp(\theta(x+\phi))-1$ and $(\exp(\theta\phi)+1)_+ = \exp(\theta\phi)+1$, so that $(\exp(\theta(x+\phi))-1)_+ + (\exp(\theta\phi)+1)_+ = \exp(\theta(x+\phi)) + \exp(\theta\phi)$. For $x \le -\phi$, we have that $(\exp(\theta(x+\phi))-1)_+ = 0$, which implies that $\lim_{\delta\to+\infty} F_S(x|\theta,\phi,\delta) = 0$. Thus, we can rewrite the limit as follows:

$$\lim_{\delta\to+\infty} F_S(x|\theta,\phi,\delta) = \frac{(\exp(\theta(x+\phi))-1)_+}{\exp(\theta(x+\phi))+\exp(\theta\phi)} \tag{45}$$

$$= F_{TL}(x|\theta,\mu=0,\phi) \tag{46}$$

which is the CDF of the truncated logistic distribution at $-\phi$. □

Appendix C. SBCM gradients and Hessian matrix

This section defines the gradients and Hessian matrices of the choice probabilities logarithm with respect to the model parameters. Let us assume we observe a choice situation with a choice set C, for which the S-BCM choice probabilities are called $P_i(\boldsymbol{\beta})$, for all $i \in C$. $\boldsymbol{\beta} = (\boldsymbol{\theta}, \varphi, \delta, \lambda) \in \mathbb{R}^{K+3}$ is a vector of attributes for which the probabilities are defined, i.e. for $\varphi > 1, \delta > 0, \lambda > 0$. To avoid overloading notation, in this section, we define $\boldsymbol{\theta} \in \mathbb{R}^K$ as the vector of utility-function parameters (i.e., $V_i = \boldsymbol{\theta}^\top \mathbf{x}_i$). No element of $\boldsymbol{\theta}$ is set to one, which means that this vector contains the scale parameter θ of the smooth truncated logistic distribution.

To calculate the log-probabilities gradient (also called the score) and Hessian matrix, we want to calculate the following quantities: $\frac{\partial \ln P_i}{\partial \boldsymbol{\theta}}, \frac{\partial \ln P_i}{\partial \varphi}, \frac{\partial \ln P_i}{\partial \delta}, \frac{\partial \ln P_i}{\partial \lambda}$, as well as the double derivatives given in the Hessian matrix below. The log probabilities gradient (a vector of size $K+3$) and Hessian (a matrix of size $(K+3)\times(K+3)$) are defined as:

$$\nabla_{\boldsymbol{\beta}} \ln P_i = \frac{\partial \ln P_i}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial \ln P_i}{\partial \boldsymbol{\theta}} & \frac{\partial \ln P_i}{\partial \varphi} & \frac{\partial \ln P_i}{\partial \delta} & \frac{\partial \ln P_i}{\partial \lambda} \end{pmatrix}$$

$$\nabla^2_{\boldsymbol{\beta}} \ln P_i = \frac{\partial^2 \ln P_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} = \begin{bmatrix} \frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} & \left(\frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \varphi}\right)^\top & \left(\frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \delta}\right)^\top & \left(\frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \lambda}\right)^\top \\ \frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \varphi} & \frac{\partial^2 \ln P_i}{\partial \varphi^2} & \frac{\partial^2 \ln P_i}{\partial \varphi\partial \delta} & \frac{\partial^2 \ln P_i}{\partial \varphi\partial \lambda} \\ \frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \delta} & \frac{\partial^2 \ln P_i}{\partial \delta\partial \varphi} & \frac{\partial^2 \ln P_i}{\partial \delta^2} & \frac{\partial^2 \ln P_i}{\partial \delta\partial \lambda} \\ \frac{\partial^2 \ln P_i}{\partial \boldsymbol{\theta}\partial \lambda} & \frac{\partial^2 \ln P_i}{\partial \lambda\partial \varphi} & \frac{\partial^2 \ln P_i}{\partial \lambda\partial \delta} & \frac{\partial^2 \ln P_i}{\partial \lambda^2} \end{bmatrix}$$

Let us define the function f_i, for any $i \in C$, as:

$$f_i(\mathbf{X}|\boldsymbol{\theta}, \varphi, \lambda) = \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\theta)) - 1 \tag{47}$$

here, we assume a linear relationship between the model attributes and the cost function. We have that $\ln P_i = \ln g_\delta(f_i) - \ln\left(\sum_{j\in C} g_\delta(f_j)\right)$. The log-probabilities gradients and Hessians with respect to $\boldsymbol{\beta}$ can be expressed with the $g_\delta(f_i), \in C$ gradients and Hessians as follows:

$$\frac{\partial \ln P_i}{\partial \boldsymbol{\beta}} = \frac{1}{g_\delta(f_i)}\frac{\partial g_\delta(f_i)}{\partial \boldsymbol{\beta}} - \frac{1}{\sum_{j\in C} g_\delta(f_j)}\sum_{j\in C}\frac{\partial g_\delta(f_j)}{\partial \boldsymbol{\beta}} \tag{48}$$

$$\frac{\partial^2 \ln P_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} = \frac{g_\delta(f_i)\frac{\partial^2 g_\delta(f_i)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} - \left(\frac{\partial g_\delta(f_i)}{\partial \boldsymbol{\beta}}\right)\cdot\left(\frac{\partial g_\delta(f_i)}{\partial \boldsymbol{\beta}}\right)^\top}{g_\delta(f_i)^2} - \frac{\left(\sum_{j\in C} g_\delta(f_j)\right)\left(\sum_{j\in C}\frac{\partial^2 g_\delta(f_j)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\right) - \left(\sum_{j\in C}\frac{\partial g_\delta(f_j)}{\partial \boldsymbol{\beta}}\right)\cdot\left(\sum_{j\in C}\frac{\partial g_\delta(f_j)}{\partial \boldsymbol{\beta}}\right)^\top}{\left(\sum_{j\in C} g_\delta(f_j)\right)^2} \tag{49}$$

The partial derivatives of g_δ can be calculated using the chain rule:

$$\frac{\partial g_\delta}{\partial \boldsymbol{\beta}} = \frac{\partial g_\delta}{\partial f_i}\frac{\partial f_i}{\partial \boldsymbol{\beta}} \tag{50}$$

$$\frac{\partial^2 g_\delta}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \frac{\partial^2 g_\delta}{\partial f_i^2} \left(\frac{\partial f_i}{\partial \boldsymbol{\beta}} \right)^2 + \frac{\partial^2 f_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \frac{\partial g_\delta}{\partial f_i} \tag{51}$$

The first and second-order derivatives of g_δ are given by:

$$\frac{\partial g_\delta}{\partial z} = \begin{cases} \left(1 + \frac{1}{\delta z}\right) \exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

$$\frac{\partial^2 g_\delta}{\partial z^2} = \begin{cases} \frac{1}{\delta^2 z^3} \exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

$$\frac{\partial g_\delta}{\partial \delta} = \begin{cases} \frac{1}{\delta^2} \exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{54}$$

$$\frac{\partial^2 g_\delta}{\partial \delta^2} = \begin{cases} \left(-\frac{2}{\delta^3} + \frac{1}{\delta^4 z}\right) \exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{55}$$

$$\frac{\partial^2 g_\delta}{\partial \delta \partial z} = \begin{cases} \frac{1}{\delta^3 z^2} \exp\left(-\frac{1}{\delta z}\right) & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

f_i has the following partial derivatives:

$$\frac{\partial f_i}{\partial \boldsymbol{\theta}} = \left(\mathbf{x}_i - \varphi \frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} \right) \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\boldsymbol{\theta})) \tag{57}$$

$$\frac{\partial f_i}{\partial \varphi} = -m_\lambda(\mathbf{X}|\boldsymbol{\theta}) \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\boldsymbol{\theta})) \tag{58}$$

$$\frac{\partial^2 f_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \left[\left(\mathbf{x}_i - \varphi \frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} \right) \left(\mathbf{x}_i - \varphi \frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} \right)^\top - \varphi \frac{\partial^2 m_\lambda}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right] \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\boldsymbol{\theta})) \tag{59}$$

$$\frac{\partial^2 f_i}{\partial \varphi^2} = m_\lambda^2(\mathbf{X}|\boldsymbol{\theta}) \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\boldsymbol{\theta})) \tag{60}$$

$$\frac{\partial^2 f_i}{\partial \boldsymbol{\theta} \partial \varphi} = -\left[m_\lambda(\mathbf{X}|\boldsymbol{\theta}) \left(\mathbf{x}_i - \varphi \frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} \right) + \frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} \right] \exp(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi m_\lambda(\mathbf{X}|\boldsymbol{\theta})) \tag{61}$$

For a vector $\mathbf{u} = (u_1 \cdots u_N)$. The partial derivatives of $m_\lambda(\mathbf{u})$ with respect to the vector components are given by:

$$\frac{\partial m_\lambda}{\partial u_i} = \frac{\exp(\lambda u_i)}{\sum_k \exp(\lambda u_k)} \left(1 + \lambda(u_i - m_\lambda(\mathbf{u}))\right) \tag{62}$$

$$\frac{\partial^2 m_\lambda}{\partial u_i \partial u_j} = \lambda \frac{\exp(\lambda u_i)}{\sum_k \exp(\lambda u_k)} \left[\left(\delta_{ij} - \frac{\exp(\lambda u_j)}{\sum_k \exp(\lambda u_k)} \right) (1 + \lambda(u_i - m_\lambda(\mathbf{u}))) + \left(\delta_{ij} - \frac{\partial m_\lambda}{\partial u_j} \right) \right] \tag{63}$$

We can then calculate the derivative of m_λ with respect to θ by using the chain rule and setting $\mathbf{u} = \theta^\top \mathbf{X}$, for instance, using,

$$\frac{\partial m_\lambda}{\partial \boldsymbol{\theta}} = \frac{\partial m_\lambda}{\partial \boldsymbol{u}} \frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}} = \mathbf{X} \frac{\partial m_\lambda}{\partial \boldsymbol{u}}$$

$$\frac{\partial^2 m_\lambda}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \frac{\partial^2 m_\lambda}{\partial \boldsymbol{u} \partial \boldsymbol{u}^\top} \left(\frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}} \right)^2 + \frac{\partial m_\lambda}{\partial \boldsymbol{u}} \frac{\partial^2 \mathbf{u}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$$

Let us write $m_\lambda(\mathbf{u}) = \mathbb{E}_p(\mathbf{u})$, being the expectation of $\mathbf{u}$ with respect to the Softmax weighting $\mathbf{p} = \left(\frac{e^{\lambda u_i}}{\sum_j e^{\lambda u_j}} \right)_{j \in \{1,\dots,N\}}$. Finally, we have the following partial derivatives for m_λ.

$$\frac{\partial m_\lambda}{\partial \lambda} = \mathrm{Var}_p(\mathbf{u}) \tag{64}$$

$$\frac{\partial^2 m_\lambda}{\partial \lambda^2} = \mathrm{Var}_p(\mathbf{u}^2) \tag{65}$$

where $\mathrm{Var}_p(\mathbf{u}) = \mathbb{E}_p(\mathbf{u}^2) - \mathbb{E}_p(\mathbf{u})^2$ Finally, we have the cross derivatives that are given by

$$\frac{\partial^2 m_\lambda}{\partial \lambda \partial \mathbf{u}} = \frac{\partial \mathrm{Var}_p(\mathbf{u})}{\partial \mathbf{u}} \tag{66}$$

$$= 2 \left[\mathrm{diag}(\mathbf{p})\mathbf{u} - (\mathbf{p}^\top \mathbf{u})\mathbf{p} \right] \tag{67}$$

Combining all the above equations allows for deriving the choice probabilities and their logarithm, the analytical gradients, and Hessian matrices with respect to the model attributes and parameters. In the following sections, we will present a few applications of these analytical gradients and Hessians.

Appendix D. Boltzmann approximation error

Let $\mathbf{x} = \left(x_1, \dots, x_N\right)$, we want to give an upper bound on the Boltzmann operator approximation error $\epsilon = |\max(\mathbf{x}) - m_\lambda(\mathbf{x})|$. We assume, without loss of generality, that $\mathbf{x}$ is ordered so that $x_1 \leq x_2 \leq \dots \leq x_N$. The approximation error is given by:

$$\begin{aligned}\epsilon(\lambda) &= x_N - \sum_{i=1}^{N} x_i \frac{e^{\lambda x_i}}{\sum_{j=1}^{N} e^{\lambda x_j}} \\ &= \sum_{i=1}^{N} (x_N - x_i) \frac{e^{\lambda(x_i - x_N)}}{1 + \sum_{j=1}^{N-1} e^{\lambda(x_j - x_N)}}\end{aligned}$$

Let us define $\bar{\epsilon}(\gamma) = \epsilon(1/\lambda)$, we can give the following Taylor approximation of the error ϵ:

$$\bar{\epsilon}(\gamma) = \bar{\epsilon}(0) + \gamma \frac{\partial \bar{\epsilon}}{\partial \gamma}\Big|_{\gamma=0} + o(\gamma)$$

$\epsilon(\gamma = 0) = \epsilon(\lambda = +\infty) = 0$ as the true maximum is obtained for infinite λ. Moreover, the partial derivative is given by the LogSum: $\frac{\partial \bar{\epsilon}}{\partial \gamma} = \log\left(\sum_{i=1}^{N} e^{(x_i - x_N)/\gamma}\right)$. This implies that, when λ goes to infinity:

$$\epsilon(\lambda) = \frac{1}{\lambda} \log\left(\sum_{i=1}^{N} e^{\lambda(x_i - x_N)}\right) + o\left(\frac{1}{\lambda}\right)$$

Given that $e^{\lambda(x_i - x_N)} \leq 1$ for all i, we have the following approximation error:

$$\epsilon(\lambda) = \frac{1}{\lambda} \log(N) + o\left(\frac{1}{\lambda}\right) = O\left(\frac{1}{\lambda}\right) \tag{68}$$

References

Abraham, I., Delling, D., Goldberg, A.V., Werneck, R.F., 2013. Alternative routes in road networks. ACM J. Exp. Algorithmics 18, http://dx.doi.org/10.1145/2444016.2444019.

Anderson, M.K., 2013. Behavioural Models for Route Choice of Passengers in Multimodal Public Transport Networks. DTU Transport, Kg.s Lyngby.

Asadi, K., Littman, M.L., 2017. An alternative softmax operator for reinforcement learning. In: Proceedings of the 34th International Conference on Machine Learning. PMLR, pp. 243–252, ISSN: 2640-3498.

Başar, G., Bhat, C.R., 2004. A parameterized consideration set model for airport choice: An application to the San Francisco Bay Area. Transp. Res. Part B: Methodol. 38 (10), 889–904. http://dx.doi.org/10.1016/j.trb.2004.01.001.

Başaran, G.G., Ingvardson, J.B., Nielsen, O.A., 2025. Does transit-oriented development (TOD) influence perceived safety and mode choice? J. Transp. Land Use Vol. 18, 237–267. http://dx.doi.org/10.5198/jtlu.2025.2548, No. 1 (2025).

Ben-Akiva, M.E., Boccara, B., 1995. Discrete choice models with latent choice sets. Consideration sets, Int. J. Res. Mark. Consideration sets, 12 (1), 9–24.http://dx.doi.org/10.1016/0167-8116(95)00002-J,

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Google-Books-ID: oLC6ZYPs9UoC.

Bhat, C.R., 2015. A comprehensive dwelling unit choice model accommodating psychological constructs within a search strategy for consideration set formation. Transp. Res. Part B: Methodol. 79, 161–188. http://dx.doi.org/10.1016/j.trb.2015.05.021.

Bierlaire, M., Hurtubia, R., Flötteröd, G., 2010. Analysis of implicit choice set generation using a constrained multinomial logit model. Transp. Res. Rec.: J. Transp. Res. Board 2175, http://dx.doi.org/10.3141/2175-11.

Bovy, P.H.L., 2009. On modelling route choice sets in transportation networks: A synthesis. Transp. Rev. 29 (1), 43–68. http://dx.doi.org/10.1080/01441640802078673.

Bovy, P.H.L., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: Behavioral and probabilistic foundations. Transportmetrica 3 (3), 173–189. http://dx.doi.org/10.1080/18128600708685672, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/18128600708685672.

Brathwaite, T., Walker, J.L., 2018. Asymmetric, closed-form, finite-parameter models of multinomial choice. J. Choice Model. 29, 78–112. http://dx.doi.org/10.1016/j.jocm.2018.01.002.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. ISBN: 978-0-08-042586-3, pp. 697–711, ISSN.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transp. Res. Part C: Emerg. Technol. 9 (4), 249–263. http://dx.doi.org/10.1016/S0968-090X(00)00036-X.

Castillo, E., Menéndez, J.M., Jiménez, P., Rivas, A., 2008. Closed form expressions for choice probabilities in the Weibull case. Transp. Res. Part B: Methodol. 42 (4), 373–380. http://dx.doi.org/10.1016/j.trb.2007.08.002.

Castro, M., Martínez, F., Munizaga, M.A., 2013. Estimation of a constrained multinomial logit model. Transportation 40 (3), 563–581. http://dx.doi.org/10.1007/s11116-012-9435-4.

Cazor, L., Watling, D.P., Duncan, L.C., Nielsen, O.A., Rasmussen, T.K., 2024. A novel choice model combining utility maximization and the disjunctive decision rules, application to two case studies. J. Choice Model. 52, 100510. http://dx.doi.org/10.1016/j.jocm.2024.100510.

Chikaraishi, M., Nakayama, S., 2016. Discrete choice models with q-product random utilities. Transp. Res. Part B: Methodol. 93, 576–595. http://dx.doi.org/10.1016/j.trb.2016.08.013.

Daly, A., Hess, S., de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. Emerging and Innovative Directions in Choice Modeling, Transp. Res. Part B: Methodol. Emerging and Innovative Directions in Choice Modeling, 46 (2), 333–341.http://dx.doi.org/10.1016/j.trb.2011.10.008,

Ding, Y., Veeman, M.M., Adamowicz, W.L., 2012. The influence of attribute cutoffs on consumers' choices of a functional food. Eur. Rev. Agric. Econ. 39 (5), 745–769. http://dx.doi.org/10.1093/erae/jbr067.

Dubey, S., Cats, O., Hoogendoorn, S., Bansal, P., 2022. A multinomial probit model with Choquet integral and attribute cut-offs. Transp. Res. Part B: Methodol. 158, 140–163. http://dx.doi.org/10.1016/j.trb.2022.02.007.

Duncan, L.C., Watling, D.P., Connors, R.D., Rasmussen, T.K., Nielsen, O.A., 2022. A bounded path size route choice model excluding unrealistic routes: formulation and estimation from a large-scale GPS study. Transp. A: Transp. Sci. 18 (3), 435–493. http://dx.doi.org/10.1080/23249935.2021.1872730, Publisher: Taylor & Francis _eprint:https://doi.org/10.1080/23249935.2021.1872730.

Elrod, T., Johnson, R.D., White, J., 2004. A new integrated model of noncompensatory and compensatory decision strategies. Organ. Behav. Hum. Decis. Process. 95 (1), 1–19. http://dx.doi.org/10.1016/j.obhdp.2004.06.002.

Fischer, S.M., 2020. Locally optimal routes for route choice sets. Transp. Res. Part B: Methodol. 141, 240–266. http://dx.doi.org/10.1016/j.trb.2020.09.007.

Florian, M., Fox, B., 1976. On the probabilistic origin of dial's multipath traffic assignment model. Transp. Res. 10 (5), 339–341. http://dx.doi.org/10.1016/0041-1647(76)90113-1.

Flötteröd, G., Bierlaire, M., 2013. Metropolis–Hastings sampling of paths. Transp. Res. Part B: Methodol. 48, 53–66. http://dx.doi.org/10.1016/j.trb.2012.11.002.

Fosgerau, M., Bierlaire, M., 2009. Discrete choice models with multiplicative error terms. Transp. Res. Part B: Methodol. 43 (5), 494–505. http://dx.doi.org/10.1016/j.trb.2008.10.004.

Frejinger, E., Bierlaire, M., 2010. On path generation algorithms for route choice models. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald Group Publishing Limited, pp. 307–315. http://dx.doi.org/10.1108/9781849507738-013.

Frejinger, E., Bierlaire, M., Ben-Akiva, M.E., 2009. Sampling of alternatives for route choice modeling. Transp. Res. Part B: Methodol. 43 (10), 984–994. http://dx.doi.org/10.1016/j.trb.2009.03.001.

Gilbride, T.J., Allenby, G.M., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. Mark. Sci. 23 (3), 391–406, Publisher: INFORMS.

Hallberg, M., Rasmussen, T.K., Rich, J., 2021. Modelling the impact of cycle superhighways and electric bicycles. Transp. Res. Part A: Policy Pr. 149, 397–418. http://dx.doi.org/10.1016/j.tra.2021.04.015.

Haque, M.B., Choudhury, C., Hess, S., 2019. Modelling residential location choices with implicit availability of alternatives. J. Transp. Land Use 12 (1), 597–618, Publisher: Journal of Transport and Land Use.

Harald Cramer, 1946. Mathematical methods of statistics. eng.

Horowitz, J.L., Louviere, J.J., 1995. What is the role of consideration sets in choice modeling? Consideration sets, Int. J. Res. Mark. Consideration sets, 12 (1), 39–54.http://dx.doi.org/10.1016/0167-8116(95)00004-L,

Kitthamkesorn, S., Chen, A., 2024. Stochastic user equilibrium model with a bounded perceived travel time. http://dx.doi.org/10.48550/arXiv.2402.18435, arXiv:2402.18435[econ,q-fin].

Lipovetsky, S., 2009. Linear regression with special coefficient features attained via parameterization in exponential, logistic, and multinomial–logit forms. Math. Comput. Modelling 49 (7), 1427–1435. http://dx.doi.org/10.1016/j.mcm.2008.11.013.

Łukawska, M., Cazor, L., Paulsen, M., Rasmussen, T.K., Nielsen, O.A., 2024. Revealing and reducing bias when modelling choice behaviour on imbalanced panel datasets. J. Choice Model. 50, 100471. http://dx.doi.org/10.1016/j.jocm.2024.100471.

Łukawska, M., Paulsen, M., Rasmussen, T.K., Jensen, A.F., Nielsen, O.A., 2023. A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset. Transp. Res. Part A: Policy Pr. 176, 103834. http://dx.doi.org/10.1016/j.tra.2023.103834.

Manski, C.F., 1977. The structure of random utility models. Theory and Decision 8 (3), 229–254. http://dx.doi.org/10.1007/BF00133443.

Martínez, F., Aguila, F., Hurtubia, R., 2009. The constrained multinomial logit: A semi-compensatory choice model. Transp. Res. Part B: Methodol. 43 (3), 365–377. http://dx.doi.org/10.1016/j.trb.2008.06.006.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econometrics 15 (5), 447–470. http://dx.doi.org/10.1002/1099-1255(200009/10)15:5{<}447::AID-JAE570{>}3.0.CO;2-1, _eprint:https://onlinelibrary.wiley.com/doi/pdf/10.1002/1099-1255%28200009/10%2915%3A5%3C447%3A%3AAID-JAE570%3E3.0.CO%3B2-1.

Moser, R., Raffaelli, R., 2014. Does attribute cut-off elicitation affect choice consistency? Contrasting hypothetical and real-money choice experiments. Process heuristics in choice analysis, J. Choice Model. Process heuristics in choice analysis, 11, 16–29.http://dx.doi.org/10.1016/j.jocm.2014.02.003,

Nielsen, O.A., 2004. A large scale stochastic multi-class schedule-based transit model with random coefficients. In: Wilson, N.H.M., Nuzzolo, A. (Eds.), Schedule-Based Dynamic Transit Modeling: Theory and Applications. In: Operations Research/Computer Science Interfaces Series, Springer US, Boston, MA, pp. 53–77. http://dx.doi.org/10.1007/978-1-4757-6467-3_4.

Nielsen, O.A., Eltved, M., Anderson, M.K., Prato, C.G., 2021. Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. Transp. Res. Part A: Policy Pr. 147, 76–92. http://dx.doi.org/10.1016/j.tra.2021.02.010.

Nocedal, J., Wright, S.J., 2006. Numerical Optimization. Springer Series in Operations Research and Financial Engineering, Springer New York, http://dx.doi.org/10.1007/978-0-387-40065-5.

Norets, A., 2010. Continuity and differentiability of expected value functions in dynamic discrete choice models. Quant. Econ. 1 (2), 305–322. http://dx.doi.org/10.3982/QE41.

Paleti, R., 2015. Implicit choice set generation in discrete choice models: Application to household auto ownership decisions. Transp. Res. Part B: Methodol. 80, 132–149. http://dx.doi.org/10.1016/j.trb.2015.06.015.

Pougala, J., Hillel, T., Bierlaire, M., 2021. Choice set generation for activity-based models.

Prato, C.G., 2009. Route choice modeling: Past, present and future research directions. J. Choice Model. 2 (1), 65–100. http://dx.doi.org/10.1016/S1755-5345(13)70005-8.

Prato, C.G., Halldórsdóttir, K., Nielsen, O.A., 2017. Latent lifestyle and mode choice decisions when travelling short distances. Transportation 44 (6), 1343–1363. http://dx.doi.org/10.1007/s11116-016-9703-9.

Radhakrishna Rao, C., 1945. Information and the accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37, 81–91.

Rasmussen, T.K., Duncan, L.C., Watling, D.P., Nielsen, O.A., 2024. Local detouredness: A new phenomenon for modelling route choice and traffic assignment. Transp. Res. Part B: Methodol. 190, 103052. http://dx.doi.org/10.1016/j.trb.2024.103052.

Simon, H.A., 1955. A behavioral model of rational choice. Q. J. Econ. 69 (1), 99–118. http://dx.doi.org/10.2307/1884852, Publisher: Oxford University Press.

Swait, J., 2001a. A non-compensatory choice model incorporating attribute cutoffs. Transp. Res. Part B: Methodol. 35 (10), 903–928. http://dx.doi.org/10.1016/S0191-2615(00)00030-8.

Swait, J., 2001b. Choice set generation within the generalized extreme value family of discrete choice models. Transp. Res. Part B: Methodol. 35 (7), 643–666. http://dx.doi.org/10.1016/S0191-2615(00)00029-1.

Swait, J., Ben-Akiva, M.E., 1987. Incorporating random constraints in discrete models of choice set generation. Transp. Res. Part B: Methodol. 21 (2), 91–102. http://dx.doi.org/10.1016/0191-2615(87)90009-9.

Swait, J., Erdem, T., 2007. Brand effects on choice and choice set formation under uncertainty. Mark. Sci. 26 (5), 679–697. http://dx.doi.org/10.1287/mksc.1060.0260, Publisher: INFORMS.

Tan, H., Xu, X., Chen, A., 2024. On endogenously distinguishing inactive paths in stochastic user equilibrium: A convex programming approach with a truncated path choice model. Transp. Res. Part B: Methodol. 183, 102940. http://dx.doi.org/10.1016/j.trb.2024.102940.

Thill, J.-C., 1992. Choice set formation for destination choice modelling. Prog. Hum. Geogr. 16 (3), 361–382. http://dx.doi.org/10.1177/030913259201600303, Publisher: SAGE Publications Ltd.

Truong, T.D., Adamowicz, W.L.V., Boxall, P.C., 2015. Modeling non-compensatory preferences in environmental valuation. Resour. Energy Econ. 39, 89–107. http://dx.doi.org/10.1016/j.reseneeco.2014.12.001.

Tsoleridis, P., Choudhury, C.F., Hess, S., 2023. Probabilistic choice set formation incorporating activity spaces into the context of mode and destination choice modelling. J. Transp. Geogr. 108, 103567. http://dx.doi.org/10.1016/j.jtrangeo.2023.103567.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: Part I – Model formulations under alternative distributions and restrictions. Transp. Res. Part B: Methodol. 77, 166–181. http://dx.doi.org/10.1016/j.trb.2015.03.008.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2018. Stochastic user equilibrium with a bounded choice model. Transp. Res. Part B: Methodol. 114, 254–280. http://dx.doi.org/10.1016/j.trb.2018.05.004.

Williams, H.C.W.L., Ortuzar, J.D., 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. Transp. Res. Part B: Methodol. 16 (3), 167–219. http://dx.doi.org/10.1016/0191-2615(82)90024-8.

Yao, R., Bekhor, S., 2022. A variational autoencoder approach for choice set generation and implicit perception of alternatives in choice modeling. Transp. Res. Part B: Methodol. 158, 273–294. http://dx.doi.org/10.1016/j.trb.2022.02.015.

Zhang, J., Timmermans, H., Borgers, A., Wang, D., 2004. Modeling traveler choice behavior using the concepts of relative utility and relative interest. Transp. Res. Part B: Methodol. 38 (3), 215–234, Publisher: Elsevier.