A doubly-differentiable bounded choice model: formulation and application to three large-scale case studies

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

[*] *Corresponding author (lauca@dtu.dk)*
[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*

January 17, 2024

Short summary

Most choice models assign non-zero probabilities to all alternatives. However, decision-makers may not consider many alternatives due to their high cost. The Bounded Choice Model (BCM) accounts for this by assigning zero probabilities to alternatives with costs exceeding some bound, thus determining a subset of alternatives individuals consider using a consistent criterion with the choice from this consideration set. The BCM is, however, non-differentiable, which prevents calculating parameter estimates' standard errors.

In this paper, we develop a doubly differentiable BCM, the $\mathcal{C}^2$ BCM. Likelihood derivatives and Hessian matrices of the $\mathcal{C}^2$ BCM are derived analytically, enabling the calculation of the model estimates' covariance matrix and elasticities. The $\mathcal{C}^2$ BCM is estimated and benchmarked with the Multinomial Logit and BCM in large-scale mode choice and route choice case studies. The $\mathcal{C}^2$ BCM provides a richer interpretation and analysis than the MNL and BCM while providing the best fit in both datasets.

Keywords: Bounded Choice Model, Consideration set, Differentiability, Analytical Hessian, Mode choice, Route choice

1 Introduction

Probabilistic choice models are used to understand and forecast people's choice behaviour and are an important tool in transport studies, e.g. when planning infrastructure improvements. Typically, these models assume that individuals choose among a *choice set* $\mathcal{C}$ of alternatives. Modellers' assumptions about individual choices are the following:

1. They choose from the whole choice set $\mathcal{C}$, and apply a choice probability model $\mathbb{P}(i|\mathcal{C})$ to determine the probability of choosing each alternative from this set, up to deterministic availability constraints. This approach may be prone to misspecification and lead to inconsistent parameter estimates (see, e.g., Swait & Ben-Akiva (1987)).
2. They choose from a subset of the choice set (usually referred to as the *consideration set*) and apply a choice probability model to this subset only. The choice probabilities of an alternative $i \in \mathcal{C}$ are then computed over all the subsets of $\mathcal{C}$, as $\sum_{X\subseteq\mathcal{C}} \mathbb{P}(X)\mathbb{P}(i|X)$, using Manski (1977) framework (e.g., Swait & Ben-Akiva (1987)). While these models assume different criteria rule consideration set formation and choice from the consideration set, Horowitz & Louviere (1995) found that the same preferences drove these two stages. Moreover, Manski's approach is computationally expensive and tractable for only small choice sets.
3. They choose from a subset of the representative universal choice set, but the consideration subset of alternatives is determined implicitly through the computation of the choice proba-

bilities from the probability model. This approach allows consistency between consideration set formation and choice from the consideration set.

While promising in many applications, this last category is yet to be investigated in a general choice modelling context. The Bounded Choice Model (BCM, Watling et al. (2018)) belongs to this category. It implicitly generates subsets of the representative universal choice sets, allocating zero probabilities to alternatives whose cost is not within some bound of a reference- alternative. This bound is estimated on observed choices (Duncan et al. (2021)). The BCM has yet only been applied to route choice (e.g., Watling et al. (2018), Duncan et al. (2023)) but is relevant in many choice contexts.

However, the BCM choice probabilities are non-differentiable. It is thus impossible to calculate standard errors and covariances of estimated parameters analytically. Moreover, proofs of the asymptotic normality of the Maximum Likelihood (ML) estimator assume the likelihood is differentiable (Norets (2010)). Non-differentiability also prevents calculating elasticities or confidence intervals for, e.g. Marginal Rates of Substitution (MRS) (Daly et al. (2012)).

To resolve the non-differentiability issue, we develop a doubly differentiable version of the BCM, the $\mathcal{C}^2$ BCM. Importantly, the $\mathcal{C}^2$ BCM maintains the core features of the BCM. This includes maintaining the same bounding, continuity, and collapsing to MNL properties. Upon formulation of the model, we derive the choice probability and likelihood gradients and Hessian matrices with respect to the model parameters and attributes. Consequently, we derive standard errors and elasticities of ML estimates. To explore these new metrics, we estimate the $\mathcal{C}^2$ BCM, and benchmark it with MNL and the original BCM, in three large-scale case studies:

- A mode choice case study on more than 20,000 observed trips in the Greater Copenhagen area. This case study allows the calculation of the demand elasticities and the analysis of the generated consideration sets.
- A bicycle route choice case study in the Greater Copenhagen area. We propose a methodology to account consistently for route overlap. We calculate the MRS for distance and their confidence intervals.
- A Public Transport route choice case study in Copenhagen area. We calculate the MRS for bus In-Vehicle Travel time (IVT) and their confidence intervals.

In all case studies, we find that the $\mathcal{C}^2$ BCM fits the data better than the BCM and MNL, and provides an enriched model interpretation.

2 Methods

The Bounded Choice Model

We assume that a decision-maker faces a choice situation with positive disutilities, which we will refer to as *cost* or *generalised cost*. We define the choice set of alternatives as $\mathcal{C}$. Each alternative i is described by K attributes that can be stored in a vector $\mathbf{x}_i = (x_{i1} \cdots x_{iK})$. For $i \in \mathcal{C}$, we call y_i the choice dummy ($y_i = 1$ is the event "the decision-maker chooses $i \in \mathcal{C}$"). The BCM from Watling et al. (2018) has the following choice probabilities:

$$\mathbb{P}(y_i = 1) = P_i^{\text{BCM}}(\mathbf{X}; \boldsymbol{\theta}, \varphi) = \frac{\left(\exp(-(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi \min_{l \in \mathcal{C}} \boldsymbol{\theta}^\top \mathbf{x}_l)) - 1\right)_+}{\sum_{j \in \mathcal{C}} \left(\exp(-(\boldsymbol{\theta}^\top \mathbf{x}_j - \varphi \min_{l \in \mathcal{C}} \boldsymbol{\theta}^\top \mathbf{x}_l)) - 1\right)_+} \tag{1}$$

$\mathbf{X} \in \mathbb{R}^{|\mathcal{C}| \times K}$ is the matrix of the $\mathbf{x}_i$'s for all $i \in \mathcal{C}$, $\boldsymbol{\theta}$ is a vector of size K of the cost function parameters so that $\mathbf{c} = \boldsymbol{\theta}^\top \mathbf{X} \in \mathbb{R}_+^{|\mathcal{C}|}$ is the vector of positive costs of the alternatives of the choice set. $\varphi \in]1, +\infty[$ is the relative surplus cost bound parameter. The function $(.)_+ = \max(0, .)$ ensures that, if $\boldsymbol{\theta}^\top \mathbf{x}_i \leq \varphi \min_{l \in \mathcal{C}} \boldsymbol{\theta}^\top \mathbf{x}_l$, i.e., if an alternative cost is higher than the relative surplus travel cost bound, the choice probabilities of the alternative i, $P_i = 0$. $\boldsymbol{\theta}$ and φ can be estimated using a MLE procedure (see Duncan et al. (2021)) for a given dataset of observed choices.

This model is continuous with respect to all the variables. However, the $(.)_+$ function is non-differentiable with respect to any parameter when one alternative reaches the bound. Similarly, the definition of the reference alternative $\boldsymbol{\theta} \rightarrow \min_{j \in \mathcal{C}} \boldsymbol{\theta}^\top \mathbf{x}_j$ is not differentiable with respect to $\boldsymbol{\theta}$. Indeed, when the index of the minimum-cost alternative $j(\boldsymbol{\theta})$ changes, the partial derivative w.r.t $\boldsymbol{\theta}$ is non-continuous.

The $\mathcal{C}^2$ Bounded Choice Model

We derive a new version of the BCM probability relation that respects the following properties:

- The choice probabilities are doubly differentiable with respect to the cost vector $\boldsymbol{\theta}$, the bound φ, and the attributes $\mathbf{X}$ on the whole domain they are defined.
- The choice probabilities tend to the MNL choice probabilities when φ tends to $+\infty$.
- The choice probability of an alternative is equal to 0 if and only if the generalized cost of this alternative is greater than φ times as much as the minimum cost among the choice set.

We are looking for a function $g_i(\mathbf{X}; \boldsymbol{\theta}, \varphi)$ such that $P_i^{\mathcal{C}^2\text{BCM}}$ respects the above-mentioned properties with

$$P_i^{\mathcal{C}^2\text{BCM}} = \frac{g_i(\mathbf{X}; \boldsymbol{\theta}, \varphi)}{\sum_{j \in \mathcal{C}} g_j(\mathbf{X}; \boldsymbol{\theta}, \varphi)}$$

We solve the two differentiability issues highlighted in the BCM definition, namely:

The non-differentiability of the $(.)_+$ function: While the function $t \rightarrow (\exp(t) - 1)_+$ is not differentiable at 0, the function $h(t) = \left(\exp(t) - \frac{t^2}{2} - t - 1\right)_+$ is doubly differentiable at the point $t = 0$. Indeed, we can see that:

$$h'(t) = (\exp(t) - t - 1)_+$$

Which is a continuous function with $g'(0) = 0$.

$$h''(t) = (\exp(t) - 1)_+$$

Which is a continuous function with $g''(0) = 0$.

The non-differentiability of the min function: To solve the non-differentiability of the min function, it is possible to use a smooth approximation to the maximum function. One example is the Mellowmax operator (Asadi & Littman (2017)), defined as:

$$\mathcal{M}_\alpha(\boldsymbol{\theta}^\top \mathbf{X}) = \frac{1}{\alpha} \ln \left(\frac{1}{|\mathcal{C}|} \sum_{j \in \mathcal{C}} e^{\alpha \boldsymbol{\theta}^\top \mathbf{x}_j} \right) \tag{2}$$

α is a fixed hyperparameter. The MellowMax function approximates the min function when $\alpha \rightarrow -\infty$. As it is infinitely differentiable for any finite value of α, we use it to replace the min function in the reference alternative definition. This leads to the following definition of g_i:

$$g_i(\mathbf{X}; \boldsymbol{\theta}, \varphi) = \left(\exp(-(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi \mathcal{M}_\alpha(\boldsymbol{\theta}^\top \mathbf{X}))) - \frac{(\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi \mathcal{M}_\alpha(\boldsymbol{\theta}^\top \mathbf{X}))^2}{2} + (\boldsymbol{\theta}^\top \mathbf{x}_i - \varphi \mathcal{M}_\alpha(\boldsymbol{\theta}^\top \mathbf{X})) - 1 \right)_+ \tag{3}$$

and hence the $\mathcal{C}^2$ BCM choice probabilities, defined as $\frac{g_i}{\sum_{j \in \mathcal{C}} g_j}$.

3 Case studies

Case study 1: Mode choice

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.

Model specification

The estimated models have the following utility specification:

$$\begin{aligned}
V_{car} &= \theta_{\text{GTT},car} \times \text{GTT}_{car} \\
V_{PT} &= \text{ASC}_{PT} + \theta_{\text{GTT},PT} \times \text{GTT}_{PT} + \theta_{acc} \times \text{Acc} + \theta_{egr} \times \text{Egr} \\
V_{cycle} &= \text{ASC}_{cycle} + \theta_{\text{GTT},cycle} \times \text{GTT}_{cycle} \\
V_{car} &= \text{ASC}_{walk} + \theta_{\text{GTT},walk} \times \text{GTT}_{walk}
\end{aligned}$$

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

$$\begin{aligned}
\text{GTT}_{car} &= \text{TT}_{car,free} + \theta_{congested} \times \text{TT}_{car,congested} + \text{TC}_{car}/VOT \\
\text{GTT}_{PT} &= \text{TT}_{inv} + \theta_{transfers} \times \text{N}_{transfers} + \theta_{wait} \times \text{WaitT} + \theta_{walk} \times \text{WalkT} + \text{TC}_{PT}/VOT \\
\text{GTT}_{cycle} &= \text{TT}_{cycle,free} + \theta_{congested} \times \text{TT}_{cycle,congested} \\
\text{GTT}_{walk} &= \text{TT}_{walk}
\end{aligned}$$

Table 1 describes the variables and fixed coefficients. $\theta_{transfers}$ is extracted from Nielsen et al. (2021), while the $\theta_{congested}, \theta_{wait}, \theta_{walk}$ values are from Hallberg et al. (2021). The car travel cost per kilometre and the Value of Time (VOT) are extracted from the Danish transport ministry[1].

Variables	Description	Constants	Value
$\text{TT}_{car,free}$	Car travel time under free flow conditions	$\theta_{congested}$	1.5
$\text{TT}_{car,congested}$	Car travel time under congested conditions	VOT	92 DKK/hour
TC_{car}	Car travel cost (car distance times 1.477DKK/km)	$\theta_{transfers}$	9
TT_{inv}	Public transport IVT	θ_{wait}	1.5
$\text{N}_{transfers}$	Public transport number of transfers	θ_{walk}	1.5
WaitT	Public transport transfer waiting time		
WalkT	Public transport transfer walking time		
TC_{PT}	Public transport travel cost		
$\text{TT}_{cycle,free}$	Cycling travel time under free flow conditions		
$\text{TT}_{cycle,congested}$	Cycling travel time under congested conditions		
TT_{walk}	Walking travel time		

Tab. 1: Variables and constants descriptions

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.

Results

The estimation results are given in Table 2. All the model estimates make intuitive sense regarding sign and relative magnitude. The BCM fits the data clearly better than the MNL, and the $\mathcal{C}^2$ BCM fits the data marginally better than the BCM.

The BCM and the $\mathcal{C}^2$ BCM bounds respectively cut out 28.7% and 21.8% of the available mode choice alternatives. Table 3 summarizes the proportion of alternatives cut out by these relative cost bounds. It shows that these cut-offs are mainly composed of walking trips, which are often too long to be considered.

We can analyse the excluded alternatives from individuals' consideration set by the $\mathcal{C}^2$ BCM bound (see Table 4). The excluded walking and cycling alternatives are mostly overly long. In contrast, the excluded Public Transport trips are rather short in areas that lack connections, where these trips are largely outperformed by car and cycling.

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

	MNL	**BCM**	**$\mathcal{C}^2$ BCM**
Cost parameters ($\boldsymbol{\theta}$)			
Alternative Specific Constants			
Car	-	-	-
Public Transport	1.279 (17.31)	0.8363	0.5685 (8.555)
Cycling	0.318 (5.681)	0.0162	-0.1255 (-2.517)
Walk	1.052 (3.671)	-0.7499	-0.7336 (-4.472)
Generalised Travel Time			
Car	0.0899 (33.19)	0.0619	0.0551 (31.26)
Public Transport	0.0197 (17.06)	0.0162	0.0166 (23.63)
Cycling	0.1009 (43.90)	0.0852	0.0815 (53.44)
Walk	0.0916 (14.04)	0.1034	0.0957 (22.39)
Public Transport variables			
Access Time	0.0972 (15.04)	0.0781	0.0770 (14.57)
Egress Time	0.0828 (16.17)	0.0652	0.0650 (15.42)
Bound (φ)	-	4.369	6.099 (39.62)
Final LL	**-11,097**	**-10,783**	**-10,781**
Adj. ρ^2	**0.579**	**0.591**	**0.591**
Number of parameters	**9**	**10**	**10**
Alternatives cut by bound	**0%**	**28.7%**	**21.8%**

Tab. 2: Model estimates, standard errors are between brackets. All parameters except the starred one are significant at the 0.01 level for the MNL and the $\mathcal{C}^2$ BCM.

	Car trips	Public Transport trips	Cycling trips	Walking trips
BCM	0%	16.8%	8.67%	77.2%
$\mathcal{C}^2$ BCM	0%	6.27%	1.10%	69.4%

Tab. 3: Percentage of available alternatives cut out by each model bound

Excluded 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

Tab. 4: Mean attributes for the included and excluded alternatives by the $\mathcal{C}^2$ BCM bound

Aggregate elasticities

We calculate the aggregate elasticities using the probability gradients. For instance, we can calculate how much, on average, a decrease in Public Transport IVT will affect the choice probabilities of all the transport modes. The elasticities output by the MNL are given in Table 5 and the $\mathcal{C}^2$ BCM elasticities are given in Table 6.

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**

Tab. 5: Aggregate point elasticities output by the MNL. The bold cells present direct elasticities, and the other ones are cross-elasticities

We observe that relative changes in Generalised Travel Time (GTT) affect particularly the choice probabilities of slow modes (cycling and walking).

Mode	GTT_{car}	GTT_{pub}	Acc	Egr	GTT_{cycle}	GTT_{walk}
Car	**-0.392**	0.137	0.061	0.059	0.231	0.0028
Public Transport	0.635	**-0.584**	**-0.296**	**-0.279**	0.515	0.072
Cycle	0.489	0.205	0.126	0.115	**-1.326**	0.113
Walk	0.185	0.254	0.201	0.186	1.703	**-4.546**

Tab. 6: Aggregate point elasticities output by the $\mathcal{C}^2$ BCM. The bold cells present direct elasticities, and the other ones are cross-elasticities

Differences can be found between the two models. For instance, the elasticity of car probabilities with respect to walking travel time is around four times smaller according to the $\mathcal{C}^2$ BCM than the MNL. This is likely because most chosen car trips (11,700 times over 12,363 car choices) are too long for walking to be considered. We also observe that a marginal increase in cycling travel time has a greater impact on the walking predicted share according to the $\mathcal{C}^2$ BCM than the MNL, suggesting a large substitution of cycling trips to walking trips.

Case study 2: Bicycle Route choice

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

The Data

The case utilised a large-scale crowd-sourced dataset of bicycle GPS trajectories received from Hövding (see Łukawska et al. (2023) for a description). The final dataset for model estimation consists of a subset of this dataset containing 4,134 trips made by 4,134 cyclists. The cyclable network can be modelled as a directed graph $G = (V, E)$ where E is the set of links and V is the set of nodes. The network size is large, with $|E| = 420,973$ and $|V| = 324,492$. The attributes of link $a \in E$ 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
- 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 the cost attribute vector $\mathbf{t}_a = (L_a, E_a, \text{No}_a, S_a, W_a)$. For a route i using a set of links $A_i \subseteq E$, these attributes are link-additive, so that the vector of cost attributes of route i is defined as $\mathbf{x}_i = \sum_{a \in A_i} x_a$.

Correlation between alternatives: The $\mathcal{C}^2$ Bounded Path Size model

Due to the complex overlapping nature of road networks, the correlation between routes (i.e. through link-sharing) should be accounted for (Florian & Fox (1976)). However, the BCM, and thus the $\mathcal{C}^2$ BCM, do not account for route overlap. Path-Size correction models (Ben-Akiva & Bierlaire (1999)) is a branch of models that account for overlap by penalizing the cost of overlapping routes. Extending the BCM to account for such, Duncan et al. (2021) recently developed the Bounded Path Size (BPS) model. The model's key feature is that it can capture correlations between only the routes with costs below the bound. Analogously modifying the BPS model to how we modified the BCM to formulate the $\mathcal{C}^2$ BCM, we formulate a doubly-differentiable $\mathcal{C}^2$ BPS model.

$$P_i^{\mathcal{C}^2\text{BPS}}(\mathbf{X};\boldsymbol{\theta},\varphi,\eta) = \begin{cases} \dfrac{\left(\gamma_i^{\mathcal{C}^2\text{BPS}}\right)^{\eta} g_j(\mathbf{X};\boldsymbol{\theta},\varphi)}{\sum\limits_{j\in\mathcal{C}} \left(\gamma_j^{\mathcal{C}^2\text{BPS}}\right)^{\eta} g_j(\mathbf{X};\boldsymbol{\theta},\varphi)} & \text{if } c_i \leq \varphi \min \mathbf{c} \\ 0 & \text{otherwise} \end{cases}$$

$\gamma_i^{\mathcal{C}^2\text{BPS}}$ is the $\mathcal{C}^2$ Bounded Path-Size correction term, calculated as follows:

$$\gamma_i^{\mathcal{C}^2\text{BPS}}(\mathbf{X};\boldsymbol{\theta},\varphi) = \begin{cases} \sum_{a\in A_i} \frac{t_a}{c_i} \frac{g_j(\mathbf{X};\boldsymbol{\theta},\varphi)}{\sum_{j\in\mathcal{C}} g_j(\mathbf{X};\boldsymbol{\theta},\varphi)\delta_{aj}} & \text{if } c_i \leq \varphi \min \mathbf{c} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where g_j is given by the $\mathcal{C}^2$ BCM choice probabilities numerator (Equation 3). A_i is the set of links of route $i \in \mathcal{C}$, $t_a = \boldsymbol{\theta}^\top \mathbf{t}_a$ is the cost of link $a \in A_i$, and $\delta_{aj} = 1$ if route j uses link a and 0 otherwise.

Results

Table 7 displays the model parameter estimates for the following models: MNL, BCM, $\mathcal{C}^2$ BCM, BPS, and $\mathcal{C}^2$ BPS.

Model	**MNL**	**BCM**	$\mathcal{C}^2$ **BCM**	**BPS**	$\mathcal{C}^2$ **BPS**
Cost parameters ($\boldsymbol{\theta}$)					
Length	28.54 (64.40)	25.46	20.97 (43.52)	14.71	11.99 (34.51)
Elevation gain	0.1069 (3.026)	0.0846	0.0849 (3.974)	0.0533	0.0533 (3.786)
No Bike infrastructure	5.162 (16.56)	4.302	3.553 (18.02)	2.343	1.819 (16.18)
Non-smooth surface	5.525 (55.37)	4.326	3.894 (37.76)	2.253	1.815 (30.76)
Wrong way	9.474 (47.89)	7.875	7.116 (37.9)	3.929	3.123 (29.43)
Path-Size coefficient (η)	-	-	-	1.643	1.636 (42.49)
Bound (φ)	-	1.110	1.128 (501.4)	1.105	1.140 (502.1)
Final LL	**-11,076**	**-10,778**	**-10,644**	**-9,910**	**-9.808**
Adj. ρ^2	**0.513**	**0.526**	**0.532**	**0.541**	**0.569**
N params	**5**	**6**	**6**	**7**	**7**
Routes cut by bound	**0%**	**66.6%**	**59.2%**	**65.8%**	**41.5%**

Tab. 7: Model estimates. The t-statistic (i.e. the coefficient divided by its standard error) is given between brackets for each doubly-differentiable model. All the parameters are significant at the 0.01 level.

The $\mathcal{C}^2$ BCM outperforms the traditional BCM and the MNL in fit. The $\mathcal{C}^2$ BCM relative cost bound is estimated higher because of the smoothness of the probability function (there is no fast increase of the choice probabilities around the bound). Accounting for the correlation between routes with the BPS and $\mathcal{C}^2$ BPS leads to the largest increase in model fit compared to the MNL.

The MRSs to length are given in Table 8. These MRS can be interpreted as the relative amount of detour a cyclist (on average) will make to avoid one of the attributes included in the models (or one meter of steep elevation). It is possible to calculate the standard errors for the MRS as (Daly et al. (2012)):

$$\sigma\left(\frac{\theta_1}{\theta_2}\right) = \frac{\theta_1^2}{\theta_1^2}\left(\frac{\Omega_{11}}{\theta_1^2} + \frac{\Omega_{22}}{\theta_2^2} - 2\frac{\Omega_{12}}{\theta_1\theta_2}\right) \tag{5}$$

where $(\Omega_{ij})_{1\leq i,j\leq 2}$ is the asymptotic variance-covariance matrix of the MLE estimates (θ_1, θ_2). We can then calculate the confidence interval (CI_α) of confidence level $100(1-\alpha)\%$:

$$\text{CI}_\alpha\left(\frac{\theta_1}{\theta_2}\right) = \frac{\theta_1}{\theta_2} \pm z_{\alpha/2} \times \sigma\left(\frac{\theta_1}{\theta_2}\right) \tag{6}$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, Φ being the cumulative distribution function of the standard normal distribution.

The MRSs of the different models are similar, even if the BPS models seem to have slightly lower MRSs to length than the other models. For instance, the MNL outputs that cyclists are willing to ride 19.4% longer to avoid cycling on a non-smooth surface, while the $\mathcal{C}^2$ BPS outputs a

Model	**MNL**	**BCM**	$\mathcal{C}^2$ **BCM**	**BPS**	$\mathcal{C}^2$ **BPS**
Elevation gain	0.0037 (± 0.00357)	0.0033	0.0040 (± 0.00197)	0.0036	0.0044 (± 0.00233)
No Bike infrastructure	0.1808 (± 0.0220)	0.1689	0.1694 (± 0.0175)	0.1592	0.1517 (± 0.0167)
Non-smooth surface	0.1936 (± 0.00866)	0.1699	0.1857 (± 0.00693)	0.1531	0.1514 (± 0.00657)
Wrong way	0.3319 (± 0.0162)	0.3093	0.3393 (± 0.0142)	0.2670	0.2605 (± 0.0128)

Tab. 8: MRS to length for each model. The 95% confidence intervals are given between brackets.

15.1% MRS between these two attributes. These results suggest that shorter routes are the most overlapping ones but also the preferred ones by cyclists. The 95% confidence intervals for the MRSs have a similar width for all models, although slightly smaller for the doubly-differentiable models, suggesting a higher precision.

Case study 3: Public Transport route choice

We estimated $\mathcal{C}^2$ BCM model on a route choice model on the Greater Copenhagen Region's public transport network. The dataset includes metro, urban rail (S-train), local trains, regional trains and busses (see Nielsen et al. (2021)). Anderson (2013) collected the 4,810 observed routes as part of the Danish travel survey. These observations are separated into work-related trips (2,553 observations) and leisure trips (2,257 observations). The alternatives to the chosen route were generated using a Doubly-Stochastic method. The attributes and parameters estimates are given in Table 9. The model does not account for route overlap Nielsen et al. (2021) did not find it a significant explanatory variable. For each trip purpose, we estimated a MNL, a BCM and a $\mathcal{C}^2$ BCM.

Trip purpose	Work			Leisure		
Model	**MNL**	**BCM**	$\mathcal{C}^2$ **BCM**	**MNL**	**BCM**	$\mathcal{C}^2$ **BCM**
Cost parameters ($\boldsymbol{\theta}$)						
IVT						
Bus	0.3534 (37.06)	0.3372	0.3272 (30.59)	0.3335 (33.04)	0.3045	0.2809 (25.12)
Metro	0.1377 (6.530)	0.1288	0.1252 (6.588)	0.1000 (4.711)	0.1012	0.0921 (5.624)
Reg. and Intercity train	0.3133 (18.38)	0.2984	0.2884 (17.09)	0.3193 (15.58)	0.2937	0.2716 (14.42)
S-Train	0.2642 (21.81)	0.2500	0.2416 (19.82)	0.2363 (18.48)	0.2184	0.2016 (17.31)
Local train	0.3201 (14.14)	0.3070	0.2988 (13.42)	0.2650 (8.433)	0.2413	0.2192 (7.702)
Transfer components						
Nb of Transfers	2.904 (26.87)	2.711	2.613 (22.53)	2.985 (26.42)	2.535	2.312 (18.97)
Transfer walk time	0.2391 (8.408)	0.2228	0.2163 (8.146)	0.2253 (7.266)	0.2198	0.2055 (7.837)
Transfer wait time	0.0545 (10.32)	0.0524	0.0514 (10.25)	0.0488 (9.088)	0.0455	0.0440 (9.206)
Other components						
Access time	0.5852 (40.02)	0.5519	0.5345 (30.98)	0.5899 (38.77)	0.5256	0.4839 (25.79)
Egress time	0.5198 (38.81)	0.4908	0.4758 (30.57)	0.5088 (38.16)	0.4474	0.4145 (25.50)
Trip highest headway	0.1666 (21.15)	0.1623	0.1579 (21.49)	0.1474 (19.81)	0.1313	0.1221 (17.84)
Bound (φ)	-	1.526	1.641 (47.02)	-	1.533	1.638 (66.59)
Final LL	**-2,391**	**-2,373**	**-2,369**	**-2,623**	**-2,579**	**-2,563**
Adj. ρ^2	**0.804**	**0.805**	**0.806**	**0.745**	**0.749**	**0.751**
N params	**11**	**12**	**12**	**11**	**12**	**12**
Routes cut by bound	**0%**	**90.2%**	**86.7%**	**0%**	**88.9%**	**85.5%**

Tab. 9: Model results for Work and Leisure trips, all parameters are significant at the .001 level

All estimated parameters are significant and corroborate with Nielsen et al. (2021). The bounds cut out between 85 and 90% of the generated routes. The BCM and $\mathcal{C}^2$ BCM estimates are similar, with a higher relative cost bound for the $\mathcal{C}^2$ version. This difference may be attributed to the probability function smoothness, as choice probabilities increase slower than for the BCM around the bound. The $\mathcal{C}^2$ BCM allows the computation of standard errors and significantly improves the model fit to the data.

The marginal rates of substitution for Bus IVT are given for each model in Table 10. The 95%

Trip purpose	Work			Leisure		
Model	**MNL**	**BCM**	**$\mathcal{C}^2$ BCM**	**MNL**	**BCM**	**$\mathcal{C}^2$ BCM**
IVT						
Metro	0.389 (± 0.112)	0.382	0.382 (± 0.108)	0.300 (± 0.119)	0.332	0.331 (± 0.108)
Reg. and Intercity train	0.884 (± 0.0784)	0.885	0.881 (± 0.0817)	0.957 (± 0.103)	0.965	0.967 (± 0.105)
S-Train	0.745 (± 0.0479)	0.741	0.738 (± 0.0480)	0.708 (± 0.0549)	0.717	0.717 (± 0.0522)
Local train	0.904 (± 0.117)	0.911	0.913 (± 0.121)	0.794 (± 0.173)	0.792	0.780 (± 0.184)
Transfer components						
Nb of Transfers	8.191 (± 0.656)	8.041	7.985 (± 0.655)	8.949 (± 0.792)	8.324	8.228 (± 0.726)
Transfer walk time	0.674 (± 0.159)	0.661	0.661 (± 0.190)	0.675 (± 0.184)	0.722	0.731 (± 0.180)
Transfer wait time	0.154 (± 0.0302)	0.156	0.157 (± 0.0305)	0.145 (± 0.0325)	0.150	0.157 (± 0.0338)
Other components						
Access time	1.651 (± 0.0789)	1.636	1.633 (± 0.0778)	1.769 (± 0.0969)	1.726	1.722 (± 0.0896)
Egress time	1.466 (± 0.0736)	1.456	1.454 (± 0.0713)	1.525 (± 0.0846)	1.469	1.475 (± 0.0764)
Highest headway in trip	0.470 (± 0.0487)	0.481	0.482 (± 0.0469)	0.442 (± 0.0487)	0.431	0.435 (± 0.119)

Tab. 10: MRS for Bus IVT, the 95% confidence intervals are given between brackets

confidence intervals have been calculated for the MNL and $\mathcal{C}^2$ BCM. Every model outputs similar MRS for both purposes. People have a large preference for the metro, and the bus is the least preferred transport mode. They do not like transfers and slightly prefer egress time over access time.

4 Conclusion

The $\mathcal{C}^2$ BCM is a new choice model that improves on the BCM by making it doubly differentiable. This allows for calculating standard errors, confidence intervals and elasticities, which is impossible with the original BCM. The $\mathcal{C}^2$ BCM was found to provide better fits to the data than MNL in all the case studies, and it was also found to outperform the original BCM. This is likely due to the smoother shape of the probability function. The $\mathcal{C}^2$ BCM also has the advantage of implicitly selecting the alternatives that individuals do not consider, which is helpful in interpreting an individual's consideration set. Finally, the $\mathcal{C}^2$ BPS model, which combines the $\mathcal{C}^2$ BCM with the Bounded Path Size (BPS) route choice model, was found to outperform both MNL and $\mathcal{C}^2$ BCM in route choice cases. This highlights the importance of dealing with correlation and infeasible routes in generated choice sets.

References

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

Asadi, K., & Littman, M. L. (2017, July). An Alternative Softmax Operator for Reinforcement Learning. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 243–252). PMLR. Retrieved 2023-11-17, from https://proceedings.mlr.press/v70/asadi17a.html (ISSN: 2640-3498)

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In R. W. Hall (Ed.), *Handbook of Transportation Science* (pp. 5–33). Boston, MA: Springer US. Retrieved 2023-12-01, from https://doi.org/10.1007/978-1-4615-5203-1_2 doi: 10.1007/978-1-4615-5203-1_2

Daly, A., Hess, S., & de Jong, G. (2012, February). Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological*, *46*(2), 333–341. Retrieved 2023-11-23, from https://www.sciencedirect.com/science/article/pii/S0191261511001482 doi: 10.1016/j.trb.2011.10.008

Duncan, L. C., Watling, D. P., Connors, R. D., Rasmussen, T. K., & Nielsen, O. A. (2021). A bounded path size route choice model excluding unrealistic routes: formulation and estimation from a large-scale GPS study. *Transportmetrica A: Transport Science*. (Publisher: Taylor and Francis Ltd.) doi: 10.1080/23249935.2021.1872730

Duncan, L. C., Watling, D. P., Connors, R. D., Rasmussen, T. K., & Nielsen, O. A. (2023, February). Formulation and solution method of bounded path size stochastic user equilibrium models – consistently addressing route overlap and unrealistic routes. *Transportmetrica A: Transport Science*, *0*(0), 1–53. Retrieved 2023-04-11, from https://doi.org/10.1080/23249935.2023.2178240 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23249935.2023.2178240) doi: 10.1080/23249935.2023.2178240

Florian, M., & Fox, B. (1976, October). On the probabilistic origin of dial's multipath traffic assignment model. *Transportation Research*, *10*(5), 339–341. Retrieved 2024-01-08, from https://www.sciencedirect.com/science/article/pii/0041164776901131 doi: 10.1016/0041-1647(76)90113-1

Hallberg, M., Rasmussen, T. K., & Rich, J. (2021, July). Modelling the impact of cycle superhighways and electric bicycles. *Transportation Research Part A: Policy and Practice*, *149*, 397–418. Retrieved 2023-12-08, from https://www.sciencedirect.com/science/article/pii/S0965856421001130 doi: 10.1016/j.tra.2021.04.015

Horowitz, J. L., & Louviere, J. J. (1995, May). What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, *12*(1), 39–54. Retrieved 2023-12-13, from https://www.sciencedirect.com/science/article/pii/016781169500004L doi: 10.1016/0167-8116(95)00004-L

Manski, C. F. (1977, July). The structure of random utility models. *Theory and Decision*, *8*(3), 229–254. Retrieved 2023-12-15, from https://doi.org/10.1007/BF00133443 doi: 10.1007/BF00133443

Nielsen, O. A., Eltved, M., Anderson, M. K., & Prato, C. G. (2021, May). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice*, *147*, 76–92. Retrieved 2023-07-18, from https://www.sciencedirect.com/science/article/pii/S0965856421000392 doi: 10.1016/j.tra.2021.02.010

Norets, A. (2010). Continuity and differentiability of expected value functions in dynamic discrete choice models. *Quantitative Economics*, *1*(2), 305–322. Retrieved 2023-11-23, from https://onlinelibrary.wiley.com/doi/abs/10.3982/QE41 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE41) doi: 10.3982/QE41

Swait, J., & Ben-Akiva, M. (1987, April). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, *21*(2), 91–102. Retrieved 2023-07-11, from https://www.sciencedirect.com/science/article/pii/0191261587900099 doi: 10.1016/0191-2615(87)90009-9

Watling, D. P., Rasmussen, T. K., Prato, C. G., & Nielsen, O. A. (2018, August). Stochastic user equilibrium with a bounded choice model. *Transportation Research Part B: Methodological*, *114*, 254–280. Retrieved 2023-04-11, from https://www.sciencedirect.com/science/article/pii/S0191261517304083 doi: 10.1016/j.trb.2018.05.004

Łukawska, M., Paulsen, M., Rasmussen, T. K., Jensen, A. F., & Nielsen, O. A. (2023, October). A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset. *Transportation Research Part A: Policy and Practice*, *176*, 103834. Retrieved 2023-11-15, from https://www.sciencedirect.com/science/article/pii/S0965856423002549 doi: 10.1016/j.tra.2023.103834