Journal of Choice Modelling 52 (2024) 100510

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Research note

A novel choice model combining utility maximization and the disjunctive decision rules, application to two case studies

Laurent Cazor [a,*], David Paul Watling [b], Lawrence Christopher Duncan [a], 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:
Disjunctive decision rule
Generalised random disjunctive model
Decision rule heterogeneity
Latent class choice models
Marginal rates of substitution

ABSTRACT

Most choice models, e.g. Multinomial Logit (MNL), rely on random utility theory, which assumes that a compensatory utility maximization decision rule explains an individual's choice behaviour. Research has shown, however, that behaviour is sometimes better explained by non-compensatory decision rules. While some research has used Latent Class Choice Models (LCCMs) to account for multiple decision rules, many of them – such as the disjunctive rule – have yet to be explored. This paper formulates, estimates, and evaluates a LCCM that combines the MNL with a Generalised Random Disjunctive Model (GRDM), a new choice model we develop. Addressing deficiencies of existing disjunctive choice models, the GRDM allows for relative importance between attributes and is insensitive to irrelevant attributes. Unlike most non-compensatory models, it is tractable and incorporates random error terms for capturing unobserved heterogeneity across choice situations. The GRDM can be expressed as a Universal Logit (UL) model, which helps derive welfare metrics such as Marginal Rates of Substitution and elasticities and makes it possible to estimate the model with traditional software packages. The LCCM combining the GRDM and the MNL is estimated in two large-scale case studies: cyclists' route choice and public transport route choice. Results are compared with other relevant LCCM specifications and the individual choice models, where it is found that the MNL + GRDM LCCM provides the best fit to the data. We also interpret the fitted parameters and calculate the Marginal Rates of Substitution, which align with behavioural expectations.

1. Introduction

The decision rule is a crucial component of choice models, transforming attributes' performance into predicted choice. Most commonly, choice models have been based on the random utility theory (e.g., McFadden (1974b)), which adopts a utility maximization decision rule. This rule is *compensatory*, i.e. people's choice is driven by a trade-off of the attributes of the alternatives, aggregated in a utility function. Under this assumption, people are willing to substitute one attribute for some quantity of another one while keeping the utility (and hence the choice probability) constant.

Conversely, *non-compensatory* decision rules assume individuals consider alternatives on an attribute-by-attribute basis and that attributes are not aggregated into a single utility function. These rules align with what psychologists describe as sequential information processing when solving problems (Newell and Simon, 1972). Examples of non-compensatory decision rules are the

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

https://doi.org/10.1016/j.jocm.2024.100510
Received 15 September 2023; Received in revised form 20 March 2024; Accepted 6 August 2024
Available online 10 August 2024
1755-5345/

conjunctive and disjunctive rules (Coombs, 1964), which assume a decision-maker will consider choosing an alternative based on either conjunction of attributes (all attributes of the alternative must satisfy some criterion) or disjunctions of attributes (at least one attribute must satisfy some criterion). Alternatively, the Lexicographic rule (Tversky, 1969; Fishburn, 1974) assumes that decision-makers first rank attributes by relative importance and then choose the best-performing alternative with respect to the main attribute. If there are ties, the remaining alternatives are compared based on the second main attribute, and so on, until either no attribute or one alternative remains. The Elimination-By-Aspects (EBA) decision rule Tversky (1972) generalizes Lexicography by allowing any criterion for each attribute. Regret minimization (Bell, 1982), according to which decision-makers minimize their anticipated regret, has also been a focus of recent research in non-compensatory choice modelling.

Traditional Random Utility Models (RUMs), such as the Multinomial Logit (MNL) model, can capture some non-compensatory behaviour. To some extent, random error terms may capture non-compensatory behaviour, treating it as the analyst's lack of knowledge of individual tastes (Hess, 2012). However, some researchers attempted to capture this behaviour systematically. For instance, Swait (2001) proposed using a functional form for utilities that included cutoffs, helping the MNL to represent conjunctive and disjunctive behaviours. Mela and Lehmann (1995) used multi-linear specifications with interaction terms. More recently, choice modellers tried to incorporate other behavioural patterns, such as the satisficing principle (e.g. González-Valdés and Ortúzar (2018)) or attribute non-attendance (Swait, 2001; Hensher et al., 2005). However, many possible interactions between attributes may exist, which makes these models prone to specification errors. Moreover, as Johnson et al. (1989) highlighted in a simulation study, RUMs cannot approximate any non-compensatory decision rule, most notably when some attributes are negatively correlated. Hess et al. (2018) reviews which behavioural assumptions a RUM can or cannot reproduce.

Consequently, some researchers have shifted from the RUM framework and developed choice models using non-compensatory decision rules. For instance, Tversky (1969) designed a choice model based on lexicography, and Tversky (1972), Recker and Golob (1979) created choice models based on the EBA decision rule. Though both are some of the most straightforward decision rules, it can also be computationally expensive to compute lexicographic choice probabilities in choice situations that include ties between attributes. More recently, the paradigm of Random Regret Minimization (RRM) allowed for capturing semi-compensatory and non-compensatory behaviour (Chorus, 2010; van Cranenburgh et al., 2015). Models using the conjunctive (e.g., Jedidi and Kohli (2005)) and disjunctive rule or dominance approach (e.g., Ehrgott et al. (2015)) have also been developed. Some recent researchers presented choice models that move further from the Random Utility Theory, such as preference accumulation models (e.g., Hancock et al. (2021)), using mathematical psychology models like Decision Field Theory. The main common weakness of these non-compensatory models is the difficulty of obtaining easily interpretable welfare metrics (Hess et al., 2018) and an increase in model estimation computational complexity.

Rather than supposing that a single decision rule can explain the choice behaviour of all individuals across a series of choice situations, some research has been carried out to account for decision rule heterogeneity. For instance, Hess et al. (2012) hypothesized that individuals within a dataset may use different decision rules. They combined several decision rules and models in a Latent Class Choice Model (LCCM, Kamakura and Russell (1989)), allowing for mixing discrete parameter distributions and heterogeneous decision rules. Compared to traditional logit models, these models showed substantial improvements in fit when combining utility maximization with other non-compensatory decision rules (e.g., lexicography, regret minimization, EBA). A few other studies combined the RRM and RUM paradigms (e.g., Dey et al. (2018) for bicycle route choice, Xu et al. (2020) for car route choice) and also showed improvements in model fit compared to traditional models. However, regret minimization assumes a pairwise comparison of alternatives, meaning that these models highly depend on the choice set composition. This may be problematic in choice situations where the set of unobserved alternatives is potentially large (e.g., route choice). Moreover, according to Hancock and Hess (2021), combining RUM and RRM in a two-class LCCM will likely lead to confounding between taste and decision rule heterogeneity.

In this paper, we continue research into models accounting for multiple decision rules. We use a LCCM to combine non-compensatory and compensatory decision rules in a two-class model. We use the traditional MNL in the compensatory part and develop a new choice model based on the disjunctive decision rule for the non-compensatory part. This new choice model builds on the Random Disjunctive Model (RDM) from Ehrgott et al. (2015),[1] which, unlike most existing non-compensatory models, includes random error terms, enabling the model to capture the analyst's lack of knowledge on the decision-maker choices. It is also less computationally expensive to compute than other non-compensatory models (e.g., RRM, lexicographic or EBA models).

However, before exploring the combination of MNL and the RDM within a LCCM, there are some weaknesses of the RDM that need to be addressed. Firstly, the RDM choice probabilities are sensitive to irrelevant attributes. Secondly, one needs to test all possible combinations of attributes to find the best specification, as models using subsets of attributes are not nested. Thirdly, the model does not give relative importance to attributes, leading to a confounding effect between preference for an attribute and the error term variances. To address these weaknesses, we develop, in this paper, a new choice model which generalizes the RDM: the Generalised Random Disjunctive Model (GRDM). The GRDM parameterizes the relative contribution of each attribute to the choice probabilities. As a consequence, irrelevant attributes have zero effect on choice probabilities. Moreover, the additional parameters implicitly determine the best specification of attributes as well as solve the confounding effect. Given the challenge of interpreting non-compensatory models for welfare analysis (Hess et al., 2018), we present the newly developed GRDM as an instance of the Universal Logit model (McFadden et al., 1976). This formulation allows for deriving model interpretation metrics, such as the

[1] The authors referred to this model as the NCSUE, as they used the model for solving the Stochastic User Equilibrium problem. We gave it a more generic name, as it may apply to any choice situation.

Marginal Rates of Substitution (MRS), LogSums, and Elasticities. It also shows how the model can be estimated using software packages such as Biogeme, Apollo, etc. An estimation example is given for Biogeme using the Swissmetro dataset.

The MNL + GRDM LCCM is estimated in two large-scale case studies focused on the context of route choice. We use two Revealed Preference datasets previously reported in the literature: a bicycle route choice dataset (from Fosgerau et al. (2023)) and a public transport route choice dataset (from Nielsen et al. (2021)). Results are compared with other relevant LCCM specifications and individual choice models regarding model fit and interpretation. Additionally, we benchmarked the predictive ability of the different model combinations using a Monte Carlo cross-validation method. The posterior class membership probabilities allow for deriving the posterior distribution of MRS among cyclists and public transport users. We tested the GRDM against two counter hypotheses that, rather than capturing another decision rule, the model captured taste heterogeneity or non-linear preferences. An experiment that accounted for multiple classes of tastes under utility maximization confirmed that the disjunctive decision rule better explains a significant part of cyclists' route choices. We also compared the results of models with logarithmic and Box–Cox (Box and Cox, 1964) transformed attributes.

The structure of the paper is as follows. Section 2 introduces the disjunctive decision rule, presents current disjunctive choice models, and formulates the new GRDM and illustrative examples of its properties. Section 3 presents the disjunctive models as logit models and details a methodology for calculating the MRS and Elasticities. Section 4 motivates the need for considering multiple decision rules jointly when modelling choice behaviour. Specifically, it demonstrates how one can struggle to capture non-compensatory choice behaviour with a compensatory choice model, and vice versa. Section 5 presents the combination of multiple decision rules in a LCCM and discusses its posterior analysis. Section 6 presents the estimation work of two case studies on bicycle route choice and public transport route choice, both in the Greater Copenhagen Area. Section 7 concludes the paper.

2. Disjunctive choice models

In this section, we assume we observe an individual $n \in \{1, \dots, N\}$ facing a choice task, where $\mathscr{C}$ is the set of alternatives. An alternative $i \in \mathscr{C}$ is described by its K attributes $\mathbf{x}_i = \left(x_{i1} \cdots x_{iK}\right)$. For $i \in \mathscr{C}$, we call y_{in} the choice dummy ($y_{in} = 1$ if the individual n chooses i and 0 otherwise). First, we will present the disjunctive decision rules (Section 2.1) and how they predict choices in a deterministic framework (Section 2.2). Then, uncertainty will be introduced by including some of these decision rules in stochastic choice models (Sections 2.3 and 2.4). The traditional utility maximization decision rule assumes that each individual n has their ordinal utility function U_n of attributes and will choose the alternative that maximizes U_n. Formally, for an individual n:

$$y_{in} = 1 \iff U_n(\mathbf{x}_i) = \max_{j \in \mathscr{C}} U_n(\mathbf{x}_j)$$

Under this decision rule, decision-makers are willing to trade an attribute for a quantity of another while keeping its utility constant, making it compensatory. Analysts usually model this individual utility function as the sum of a deterministic part based on attribute performance and a stochastic part that models their lack of knowledge of the decision-maker. For an individual n, facing a choice set $\mathscr{C}$, a RUM gives the probability of the event $(y_{in} = 1)$:

$$\mathbb{P}(y_{in} = 1 \mid \mathbf{X}) = \mathbb{P}(U_n(\mathbf{x}_i) = \max_{j \in \mathscr{C}} U_n(\mathbf{x}_j))$$
$$\text{where } \; U_n(\mathbf{x}_i) = V_n(\mathbf{x}_i) + \epsilon_{in}$$

where $V_n(\mathbf{x}_i)$ is a function of the observed attributes and ϵ_{in} is the random error term for individual n and alternative i. $\mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_{|\mathscr{C}|})^\top$ is the matrix of the choice set attributes. We will now present choice models that use a disjunctive decision rule.

2.1. *The disjunctive decision rule*

The disjunctive decision rule assumes that decision-makers consider attributes separately, i.e., not aggregated in a single function. For each attribute[2] $k \in \{1, \dots, K\}$, the decision maker has a criterion χ_k that the attribute can pass or fail. Let us write $x_{ik} \sim \chi_k$ if alternative i passes χ_k. Examples of criteria can be, e.g., being one of the best alternatives among the choice set, being better than a certain threshold, not being much worse than a reference alternative, etc. The flexibility in the criteria means that an infinite number of decision rules can be invented and included in choice models.

Under a disjunctive decision rule, the analyst assumes that the decision maker will select any alternative that respects the criterion for at least one attribute.

$$y_{in} = 1 \iff \exists k \in \{1, \dots, K\} \text{ s.t. } x_{ik} \sim \chi_k$$

In the example below, let us consider a three-alternative choice situation ($\mathscr{C} = \{1, 2, 3\}$). These alternatives have two attributes: travel time (TT) and travel cost (TC). Alternative 1 is fast but expensive; Alternative 2 is cheap but long, and Alternative 3 is a compromise regarding travel time and cost. The attribute values are given in Table 1.

Under a disjunctive decision rule, for each attribute's criterion is *being the best among the choice set*, the decision-maker may choose Alternative 1 or 2. To predict a choice between the two, we must then use another decision rule, which could be, e.g., changing the criterion to something more stringent, assuming a random choice between the two alternatives, or relying on utility maximization to

[2] Attributes can also be grouped into sets of attributes.

Table 1
Example of a choice situation with three alternatives and two attributes.

	TT	TC
Alternative 1	1	2
Alternative 2	2	1
Alternative 3	1.2	1.2

model the choice. Manski and Lerman (1977), Swait and Ben-Akiva (1987), or Ben-Akiva and Boccara (1995) developed two-stage models, on which the first stage filtered the considered alternatives to be considered with a non-compensatory decision rule, and then modelled choice between these alternatives with a compensatory model. An overview of other existing decision rules (Conjunctive, Lexicographic and EBA) and their application to this example can be found in Appendix A.

2.2. The deterministic disjunctive model (DDM)

Let us assume a choice model in which the decision-maker considers each attribute individually and uses the disjunctive decision rule, for which each attribute criterion χ_k is "being the best alternative among the choice set". We further assume the decision-maker chooses indifferently between all the alternatives that respect the criterion for at least one attribute. According to this model, the choice probability $P_i^{\mathrm{DDM}} := \mathbb{P}(y_{in} = 1 \mid \mathrm{DDM}, \mathbf{X})$ of an alternative $i \in \mathscr{C}$ (independent of an individual) can be written as:

$$P_i^{\mathrm{DDM}} = \frac{1 - \prod_{k=1}^{K}\left(1 - \frac{\mathbb{1}_{ki}}{\sum_{l\in\mathscr{C}} \mathbb{1}_{kl}}\right)}{\sum_{j\in\mathscr{C}}\left[1 - \prod_{k=1}^{K}\left(1 - \frac{\mathbb{1}_{kj}}{\sum_{l\in\mathscr{C}} \mathbb{1}_{kl}}\right)\right]} \tag{1}$$

where $\mathbb{1}_{ki} = 1$ if alternative i is among the bests in the choice set for attribute k and 0 otherwise. The quantity $\mathbb{1}_{kj}/\sum_{l\in\mathscr{C}} \mathbb{1}_{kl}$ means that, for an attribute k, if a number $C_k = \sum_{l\in\mathscr{C}} \mathbb{1}_{kl}$ of alternatives are equally the best performing in the choice set, each of these has an equal split probability $1/C_k$ of passing criterion χ_k. The model properties are further presented in an example in Section 2.6.

2.3. The random disjunctive model

The Random Disjunctive Model (RDM) (Ehrgott et al., 2015) is based on the same decision rule and criteria as the DDM but adds uncertainty to the model of the analyst's lack of knowledge of the individual's relative sensitivity to attributes change. For an alternative i and its attribute k, we define $\tilde{x}_{ik}$ by adding a random error term to x_{ik}, i.e.,

$$\tilde{x}_{ik} = x_{ik} + \epsilon_{ik}$$

If we assume that, for every $i, k, \epsilon_{ik} \sim \text{Gumbel}\left(0, 1/\alpha_k^2\right)$, where α_k parametrizes the absolute uncertainty around an attribute perceived value, and that these error terms are independent, we get the RDM choice probabilities $P_i^{\mathrm{RDM}} := \mathbb{P}(y_{in} = 1 \mid \mathrm{RDM}, \boldsymbol{\alpha}, \mathbf{X})$ for an alternative $i \in \mathscr{C}$:

$$P_i^{\mathrm{RDM}} = \frac{1 - \prod_{k=1}^{K}\left(1 - \frac{\exp(\alpha_k x_{ik})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right)}{\sum_{l\in\mathscr{C}}\left[1 - \prod_{k=1}^{K}\left(1 - \frac{\exp(\alpha_k x_{lk})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right)\right]} \tag{2}$$

$\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_K) \in \mathbb{R}^K$ is a set of parameters to be estimated. Adding this error term also allows the model to be more flexible. Indeed, an alternative close to being the best alternative in an attribute, even if it is not the best in any attribute, will have a larger choice probability than an alternative far from this criterion. One can interpret the choice probability of an alternative as being proportional to the probability of it being chosen for at least one of its attributes, i.e.,

$$P_i^{\mathrm{RDM}} \propto 1 - \prod_{k=1}^{K}\left(1 - \mathbb{P}(\tilde{x}_{ik} \geq \max_{j\in\mathscr{C}} \tilde{x}_{jk})\right) = \mathbb{P}\left(\bigcup_{k=1}^{K}(\tilde{x}_{ik} \geq \max_{j\in\mathscr{C}} \tilde{x}_{jk})\right)$$

This model, however, has some theoretical weaknesses:

Sensitivity to irrelevant attributes: We call an attribute *irrelevant* if every alternative i from the choice set $\mathscr{C}$ has the same choice probability $1/|\mathscr{C}|$. This can happen when the attribute has the same value for each alternative or when the decision-maker is indifferent to the attribute. If we add many of these attributes, the probability of at least one attribute being perceived as the best performing will tend to be one for any alternative. If we add N irrelevant attributes to the K other attributes, then for all $i \in \mathscr{C}$:

$$\mathbb{P}\left(\bigcup_{k=1}^{K}(\tilde{x}_{ik} \geq \max_{j\in\mathscr{C}} \tilde{x}_{jk})\right) = 1 - \left(1 - \frac{1}{|\mathscr{C}|}\right)^N \prod_{k=1}^{K}\left(1 - \frac{\exp(\alpha_k x_{ik})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right) \xrightarrow[N\to+\infty]{} 1 \tag{3}$$

as $(1 - 1/|\mathscr{C}|)^N$ tends to 0 when N tends to $+\infty$. This implies that, for each alternative, the probability that the disjunctive rule is satisfied tends to be one. As a result, P_i^{RDM} tends to $1/|\mathscr{C}|$ for every $i \in \mathscr{C}$, i.e., to an equal split probability.

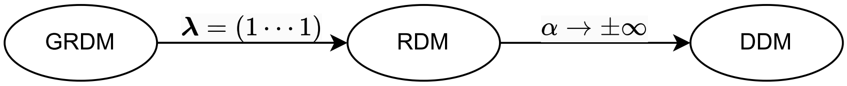

Fig. 1. Collapsing properties of the presented disjunctive models.

Model specification issues: Models using nested sets of attributes are not nested (i.e. a constrained version of each other on the parameter space). This weakness complicates the model specification, as one needs to test all the possible combinations of attributes to find the best-performing one.

Attribute relative importance: The model does not explicitly give relative importance to attributes. However, if observed choices reveal a preference for one attribute, a confounding effect may appear: the RDM model will explain this preference by giving a lower variance to the error term linked to this attribute so that the best-performing alternative will have a higher choice probability. However, it is impossible to disentangle heterogeneity (variance of the error term) and the importance of an attribute.

The novel model, developed in the following paragraphs, aims to solve these shortcomings.

2.4. The generalised random disjunctive model

To solve the weaknesses of the RDM discussed above, we develop an extended version of the RDM: the Generalised RDM (GRDM). For each attribute k, we add an exponent $\lambda_k \geq 0$ to the probability of not being chosen for this attribute. This exponent adds flexibility to the probability relation, allowing some attributes to influence the choice probabilities more than others and some of them to have no influence. The GRDM choice probabilities $P_i^{\text{GRDM}} := \mathbb{P}(y_{in} = 1 \mid \text{GRDM}, \boldsymbol{\alpha}, \lambda, \mathbf{X})$ of an alternative $i \in \mathscr{C}$ are given by:

$$P_i^{\text{GRDM}} = \frac{1 - \prod_{k=1}^{K} \left(1 - \frac{\exp(\alpha_k x_{ik})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right)^{\lambda_k}}{\sum_{l\in\mathscr{C}} \left[1 - \prod_{k=1}^{K} \left(1 - \frac{\exp(\alpha_k x_{lk})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right)^{\lambda_k}\right]} \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_K) \in \mathbb{R}^K$ and $\lambda = (\lambda_1 \cdots \lambda_K) \in \mathbb{R}_+^K$ are sets of parameters to be estimated. It is to be noted that at least one λ_k must be strictly larger than 0. Otherwise, the probabilities cannot be computed. It is important to notice that, due to the error terms structure, and unlike the DDM, the RDM and GRDM always give a non-zero probability to every alternative in the choice set. The GRDM generalises the RDM, as the RDM is a constrained version of the GRDM where all the λ_ks are set to 1.

The GRDM can be interpreted in a similar way as we interpreted the RDM in the previous subsection, by looking at the following proportionality relation:

$$P_i^{\text{GRDM}} \propto 1 - \prod_{k=1}^{K} \left(1 - \mathbb{P}(\tilde{x}_{ik} \geq \max_{j\in\mathscr{C}} \tilde{x}_{jk})\right)^{\lambda_k}$$

This relation can be interpreted as the probability of being chosen for at least one attribute with λ_k draws of the error term for each attribute k (λ_k is a real number, so one must be able to imagine a non-integer number of draws). The higher the λ_k, the more draws of the error term for a given attribute, and the higher the choice probabilities will be for this attribute, regardless of its relative performance.

This model solves the three RDM weaknesses we discussed:

Sensitivity to irrelevant attributes: An irrelevant attribute k will have its λ_k estimated to 0 and will not influence choice probabilities.

Model specification issues: The GRDM has the property that models specified with nested sets of attributes are nested into each other (a constrained version in the parameter space). This is because a model without attribute k_0 equals a model including k_0 and setting $\lambda_{k_0} = 0$. Consequently, the GRDM solves the attribute/model specification issue by implicitly specifying the best attributes (by setting attribute exponents to zero).

Attribute relative importance: A higher relative $\lambda_k > \lambda_l$ value for attributes k and l means that the performance of attribute k has a larger impact on choice probabilities than attribute l if both have the same error term variances.

2.5. Collapsing properties

Fig. 1 summarises the collapsing properties of the three presented models. The DDM is a RDM with error terms that have zero variance, which can be parametrized by setting the vector α_k to $+\infty$ if the attribute is desirable and to $-\infty$ if it is not desirable. The RDM is a GRDM with an equal contribution of each attribute to the choice probabilities and a λ_k set to 1 for each k.

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

	A	*B*	*C*
Alternative 1	1	1	2
Alternative 2	2	1	2
Alternative 3	2	2	1
Alternative 4	1.1	1.1	1.1

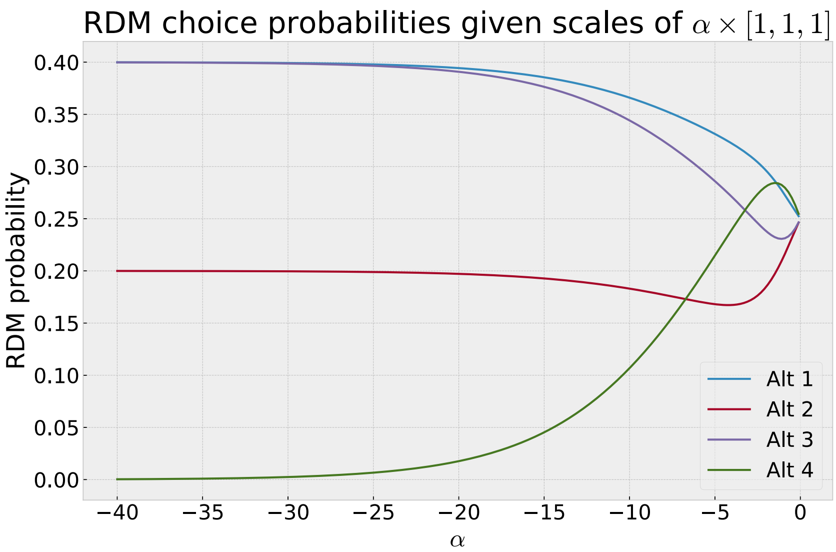

Fig. 2. Example for a RDM with equal error term variances for each attribute.

2.6. Illustrative examples

In this subsection, we will show how the three presented disjunctive models predict choice probabilities using the example from Table 2.

In this choice situation, we assume that a lower value for an attribute is better. Then, alternative 1 performs best in attributes A and B. Alternative 2 performs best in attribute B, but is tied with alternative 1. Alternative 3 performs best in attribute C, but is not tied with other alternatives. Alternative 4 is the compromise alternative, which does not perform best in any attribute.

The DDM: The DDM probabilities are given by $P_1^{\mathrm{DDM}} \propto 1-(1-1/2)(1-1)=1$, $P_2^{\mathrm{DDM}} \propto 1/2$ and $P_3^{\mathrm{DDM}} \propto 1$. This means that $P_1 = P_3 = 0.4$ and $P_2 = 0.2$. Alternatives 1 and 3 have the same choice probabilities because they perform best for one attribute without being tied, while alternative 2 performs best in one attribute but is tied. The compromise alternative probability P_4^{DDM} is equal to 0.

Random models:

We can see that the RDM probabilities (see Fig. 2) tend to the DDM probabilities when α tends to $-\infty$. For lower values of α, however, the uncertainty introduced by the error term allows the compromise alternative to have a non-zero choice probability to the point that it is the most probable alternative for very small α values (or large error term variances). To some extent, the RDM can thus also accommodate compensatory behaviour. Unlike the MNL, the RDM can reproduce the market shares even if no one chooses the compromise alternative. However, when estimating the model, if more observations showed a choice for Alternative 1 than Alternative 3, i.e., a preference for attribute A over attribute C, the model will assume that the value of α_A is more negative than the one of α_C, which can be interpreted as a lower variance in perception of travel time than travel cost ($\boldsymbol{\alpha}^2$ being inversely proportional to the error terms variance). As illustrated in Fig. 3, the GRDM can solve this issue using a more flexible functional form. Compared to the RDM, when adding attributes of relative importance, we see that the probability of alternative 1 increases, while the one of alternatives 2 and 3 decreases. However, with significant negative α values, the probability of alternatives 1 and 3 have the same limit, as they both perform best in an attribute without being tied.

Estimation example: Under the choice scenarios from Table 3, we estimate a RDM and a GRDM choice model on this dataset. The results are presented in Table 4. Let us assume we observe 1000 choices for each of three choice situations whose attributes and number of choices are given in Table 3.[3]

[3] The number of choice situations has been increased from the previous example to allow more variation in the choice set. Models estimated on only one choice situation raised identification problems.

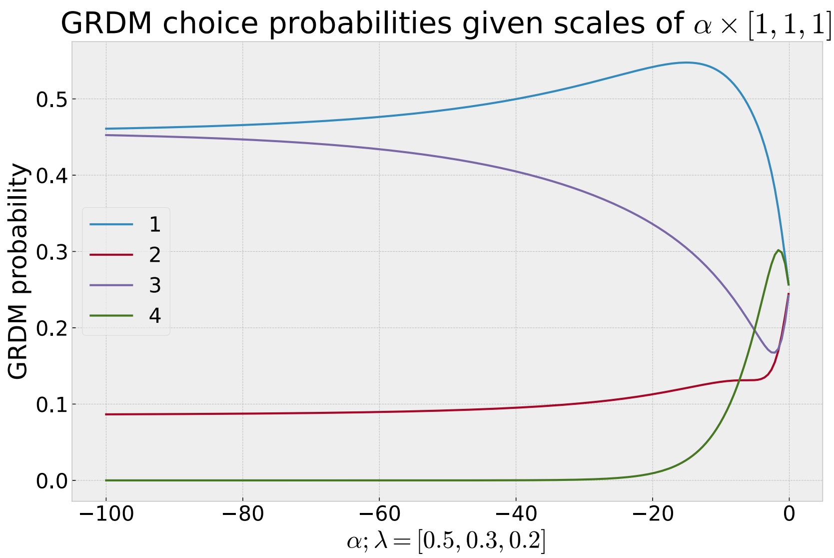

Fig. 3. Example for a GRDM with equal error term variances for each attribute.

Table 3
Example of a choice situation with three alternatives and two attributes.

Situation 1	TT	TC	Situation 2	TT	TC	Situation 3	TT	TC
Alt 1: $N_1 = 600$	1	2	Alt 1: $N_1 = 400$	3	4	Alt 1: $N_1 = 700$	2	3
Alt 2: $N_2 = 400$	2	1	Alt 2: $N_2 = 600$	4	3	Alt 2: $N_2 = 300$	3	2
Alt 3: $N_3 = 0$	1.2	1.2	Alt 3: $N_3 = 0$	3.4	3.4	Alt 3: $N_3 = 0$	2.6	2.6

Table 4
Estimation results with observed shares.

Attribute	RDM	GRDM ($\alpha\|\lambda$)	
Travel Time	−91.75	−244.9	0.081
Travel Cost	−80.16	−183.5	0.031
Final LL	**−2079.4**	**−2059.3**	
Adj. ρ^2	**0.370**	**0.376**	

The negative signs for the $\boldsymbol{\alpha}$'s make sense because they mean that choice probabilities decrease with travel costs and duration. Their relative magnitude also makes sense, as more choices are made for the alternative with the best travel time, meaning that this attribute is – in aggregate – more important than cost. The GRDM performs better than the RDM, which was to be expected as the RDM is a constrained version of the GRDM with the exponents being fixed to 1. It also shows a preference for travel time with the exponents as $\lambda_{TT} > \lambda_{TC}$. The estimated error term variances are minimal because it must be unlikely that the compromise alternative is considered the best performing in one attribute.

3. Welfare interpretation of the disjunctive models

The disjunctive models we developed can be seen as instances of the Universal Logit (UL) model, which was first described by McFadden et al. (1976) as a way to relax the MNL Independance of Irrelevant Alternatives (IIA). The UL allows the deterministic utility of an alternative to be influenced by the other alternatives. As highlighted by Hess et al. (2018), this framework may not always be consistent with RUM, and can capture non-compensatory or semi-compensatory behaviour (see e.g., Gaundry and Dagenais (1979), Chorus (2012b)). In this section, we interpret the disjunctive models using standard econometric approaches (see Dekker (2014) for similar work on the semi-compensatory Random Regret Minimization (RRM) models). We present the GRDM as a UL model, and then utilise this to derive formulas for the Marginal Rates of Substitution (MRS), LogSums and elasticities. Another key benefit of this formulation of the GRDM is that it allows the GRDM to be estimated using software packages such as Biogeme,[4] Apollo, etc., by entering the desired formulas for μ_i rather than utilities.

[4] An estimation of the RDM, GRDM and combined Latent Class MNL + RDM and MNL + GRDM has been made using the Pandas Biogeme package and the public Swissmetro dataset. The source code and data files are available on the GitHub repository: https://github.com/LauCaz/disjunctive_models.

Table 5
Example of a choice situation with three alternatives and two attributes.

	TT	TC
Alt 1	1	2
Alt 2	2	1
Alt 3	x	y

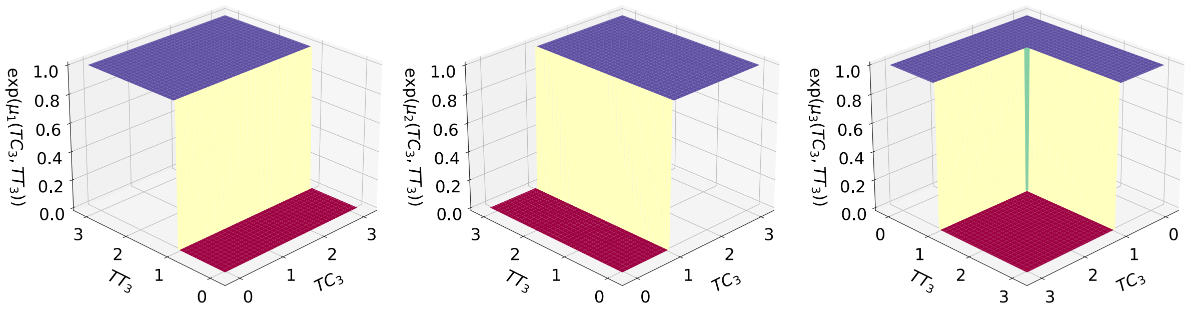

Fig. 4. $\exp(\mu_1^{\text{DDM}})$, $\exp(\mu_2^{\text{DDM}})$ and $\exp(\mu_3^{\text{DDM}})$ as a function of TT_3 and TC_3.

3.1. The disjunctive models as logit models

We can notice that the GRDM choice probabilities can be written using the logit formula:

$$P_i^{\text{GRDM}} = \frac{e^{\mu_i}}{\sum_{j\in\mathscr{C}} e^{\mu_j}} \tag{5}$$

where, for each alternative, μ_i is defined as:

$$\mu_i = \ln\left(1 - \prod_{k=1}^{K}\left(1 - \frac{\exp(\alpha_k x_{ik})}{\sum_{l\in\mathscr{C}}\exp(\alpha_k x_{jk})}\right)^{\lambda_k}\right) \tag{6}$$

We can also find the RDM (μ_i^{RDM}) and DDM (μ_i^{DDM}) using the collapsing properties from Section 2.5. The different disjunctive models can then be derived by assuming that individuals want to maximise their v_i, which is defined as:

$$v_i = \mu_i + \epsilon_i$$

Under the assumption that all the $\epsilon_i, i \in \mathscr{C}$ are iid Gumbel (0, 1) distributed, we can derive the DDM, RDM, and GRDM choice probabilities using the logit formula. The interpretation of the quantity μ_i is less straightforward than for the RUM counterpart. For the disjunctive models, $\mu_i \in]-\infty, 0[$ is the log-probability that alternative i performs best in *at least* one attribute. It is thus an indicator of its attractivity under a disjunctive decision rule. For instance, if i performs best in an attribute without being tied, its μ_i will be close to $\ln(1) = 0$. If, conversely, it is outperformed on every attribute, its μ_i will be close to $\ln(0) = -\infty$.

Examples: In the example below, let us reconsider the same choice situation as in Table 1. The only difference is that alternative 3 has variable travel time and travel cost (see Table 5).

Fig. 4 plots the $\exp(\mu_i^{\text{DDM}})$ values (we applied the exponential as some values of μ_i^{DDM} are $-\infty$). We can see from the plot that μ_i is zero if and only if an alternative performs best in one attribute and $-\infty$ otherwise, respecting the disjunctive rule.

Fig. 5 shows the $\exp(\mu_i^{\text{RDM}})$ (or $\exp(\mu_i^{\text{GRDM}})$ with $\lambda = (1, 1)$) values when we modify the travel time and travel costs of alternative 3. Unlike RUMs, modifying the attributes of an alternative also modifies the μ_i of the other alternatives, meaning that the disjunctive models do not exhibit the IIA property. We see that μ_1^{RDM} grows with TT_3, while it is not much impacted by TC_3. This is because travel cost is already outperformed by alternative 2, while travel time is not. A decrease in travel time for alternative 3 thus decreases the disjunctive attractivity of alternative 1. We see the same pattern for the value of μ_2^{RDM}, which is much more impacted by TC_3 than by TT_3. μ_3^{RDM} is high if either TT_3 or TC_3 is high, which illustrates the disjunctive decision rule. As $\alpha_{TT} = \alpha_{TC}$, we observe a symmetry of the slopes concerning travel time and cost.

Fig. 6 shows the $\exp(\mu_i^{\text{GRDM}})$ values, for which we have given relative importance for travel time that was ten times smaller than travel cost. The only difference with the previous example (Fig. 5) is that the relative importance of travel time has been reduced. This is translated by a lower gradient with respect to TT_3 for all the μ_i^{GRDM}s.

3.2 Marginal rates of substitution

While the DDM can be considered a pure non-compensatory model (i.e., there is no substitution between attributes), the RDM and GRDM allow for some compensation between attributes. They could be referred to as semi-compensatory models. In this subsection,

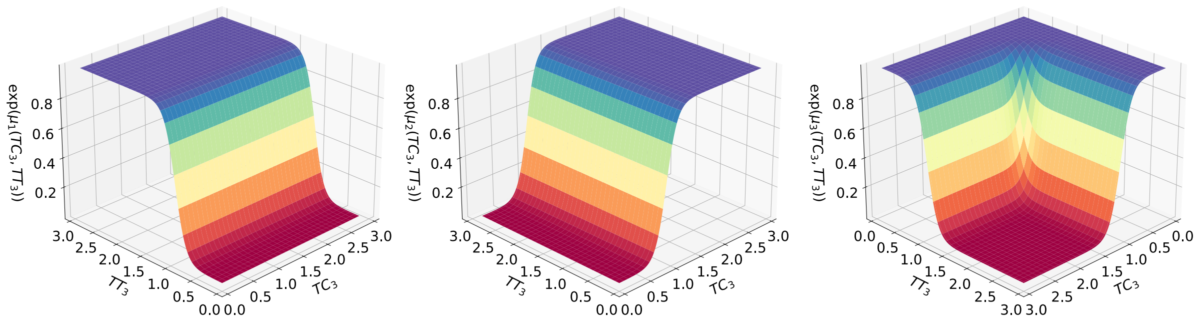

Fig. 5. $\exp(\mu_1^{\text{GRDM}})$, $\exp(\mu_2^{\text{GRDM}})$ and $\exp(\mu_3^{\text{GRDM}})$ as a function of TT_3 and TC_3, $\alpha_{TT} = \alpha_{TC} = -10$, $\lambda_{TT} = \lambda_{TC} = 1$.

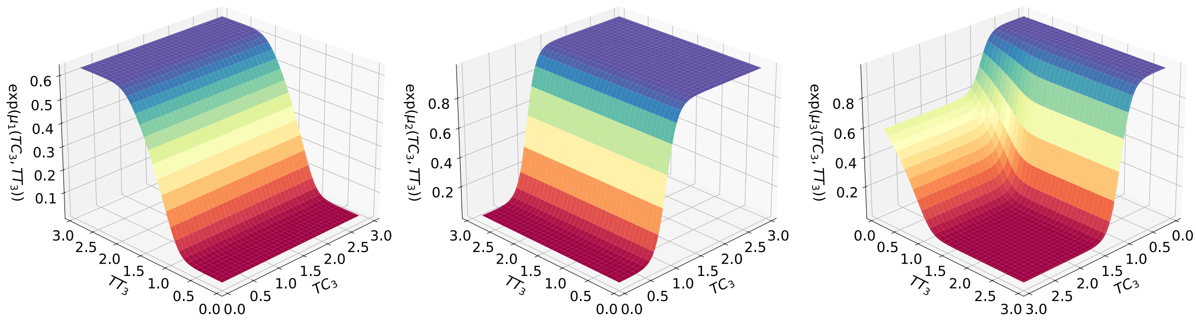

Fig. 6. $\exp(\mu_1^{\text{GRDM}})$, $\exp(\mu_2^{\text{GRDM}})$ and $\exp(\mu_3^{\text{GRDM}})$ as a function of TT_3 and TC_3, $\alpha_{TT} = \alpha_{TC} = -10$, $\lambda_{TT} = 0.1$, $\lambda_{TC} = 1$.

we thus study how attributes can compensate for each other. Following the work from Chorus (2012b), an obvious candidate for the disjunctive models MRS is, for an alternative $i \in \mathscr{C}$ and two attributes k, l:

$$MRS_{kl}^{i} = \frac{\partial \mu_i / \partial x_{ik}}{\partial \mu_i / \partial x_{il}}$$

These are not defined for the DDM (because they are either 0 or ∞). For the GRDM, they are given by (see Appendix C for a proof):

$$MRS_{kl}^{i} = \frac{\lambda_k \alpha_k P_{ik}}{\lambda_l \alpha_l P_{il}} \tag{7}$$

where $P_{ik} = \frac{\exp(\alpha_k x_{ik})}{\sum_{l \in \mathscr{C}} \exp(\alpha_k x_{jk})}$ is the choice probability of i for attribute k. These MRS are easy to interpret. For the RDM, they are linked to the ratio of sensitivities to the different attributes, multiplied by the ratio of the probabilities of i being chosen for these attributes. For the GRDM, the MRS depend both on the sensitivity to attribute change and to their relative importance. We can see from Eq. (7) that the MRS are both dependent on the attribute k and l values for alternative i, but also all the other alternatives from the choice set.

Examples: Using the same examples as in the previous subsection, we plot each alternative MRS between travel time and cost (the Value of Time, or VoT) as a function of the attributes of alternative 3. These are calculated as $\text{VoT}_i = \frac{\partial \mu_i / \partial TT_i}{\partial \mu_i / \partial TC_i}$. Fig. 7 shows the RDM VoT when we modify the travel time and costs of alternative 3. These were plotted using a logarithmic scale to better see the low VoT values. We can see that the VoT of the users of alternative 1 (cheap and slow) is always smaller than the MNL output, while it is always higher for the users of alternative 2 (expensive and fast). For both alternatives, we observe the same pattern of increasing with TT_3 and decreasing with TC_3. The higher TT_3, the more a marginal change of TT_1 (or TT_2) will impact μ_1 (or μ_2), and the same for travel costs. For the users of alternative 3, the VoT is particularly high when TC_3 is high and TT_3 is low. In that case, a marginal change of TT_3 greatly impacts μ_3, as the alternative performs best in that attribute. Conversely, a marginal change of TC_3 has a low impact on μ_3, as the alternative already performs best in another attribute. Hence, the ratio of partial derivatives is particularly high. Using the same reverse arguments, this ratio is particularly low for low values of TC_3 and high values of TT_3.

Fig. 8 plots the VoT in the case Travel Time is parametrized as a less critical attribute than travel cost. These can be comparably interpreted to the RDM. The VoT is, however, $\lambda_{TC}/\lambda_{TT} = 10$ times smaller than the RDM output, as the model considers Travel cost to be a ten times more important attribute than Travel time.

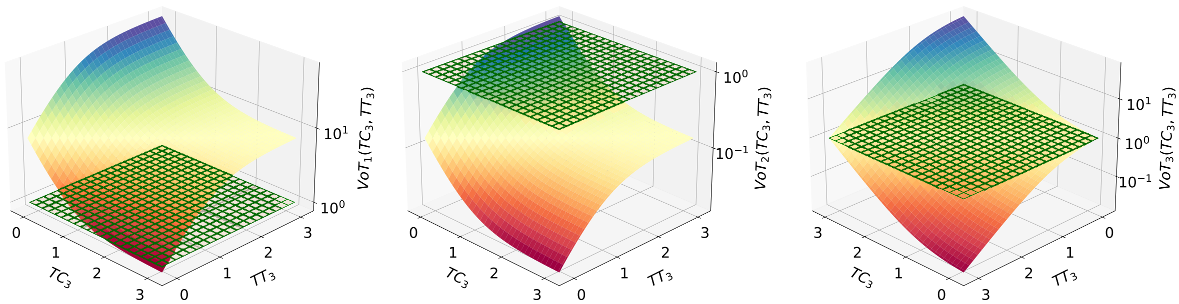

Fig. 7. $\text{VoT}_1^{\text{RDM}}$, $\text{VoT}_2^{\text{RDM}}$ and $\text{VoT}_3^{\text{RDM}}$ as a function of TT_3 and TC_3, $\alpha_{TT} = \alpha_{TC} = -2$. The green surface corresponds to the MNL VoT for the same values of α_{TT}, α_{TC}.

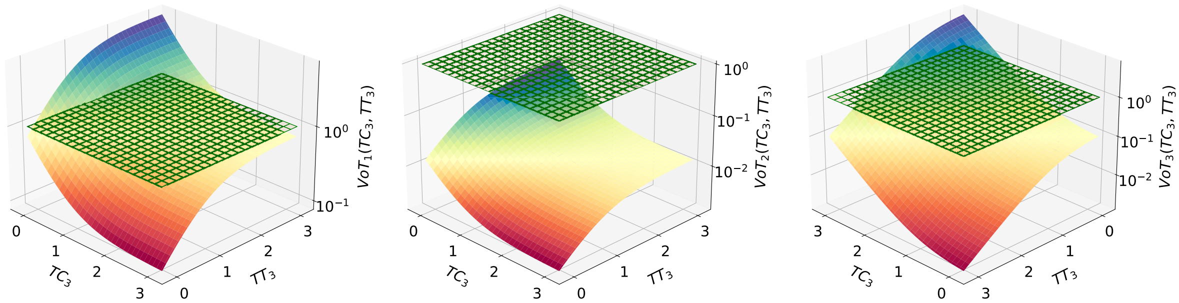

Fig. 8. $\text{VoT}_1^{\text{GRDM}}$, $\text{VoT}_2^{\text{GRDM}}$ and $\text{VoT}_3^{\text{GRDM}}$ as a function of TT_3 and TC_3, $\alpha_{TT} = \alpha_{TC} = -2$, $\lambda_{TT} = 0.1$, $\lambda_{TC} = 1$. The green surface corresponds to the MNL VoT for the same values of α_{TT}, α_{TC}.

3.3 LogSum

We can also derive a log sum metric, similarly as what Chorus (2012a) did for the RRM, defined as the expected maximum value of μ_i in a choice set $\mathscr{C}$. It is given by:

$$LS_{\mathscr{C}} = \mathbb{E}(\max_{j\in\mathscr{C}} v_j) = \ln\left(\sum_{j\in\mathscr{C}} e^{\mu_j}\right)$$

For the GRDM, it is thus given by:

$$LS_{\mathscr{C}}^{\text{GRDM}} = \ln \sum_{l\in\mathscr{C}} \left(1 - \prod_{k=1}^{K}\left(1 - \frac{\exp(\alpha_k x_{lk})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}\right)^{\lambda_k}\right) \tag{8}$$

This LogSum is particularly high when many alternatives in the choice set verify the disjunctive decision rule and particularly low when few alternatives in the choice set verify it. For instance, when one alternative dominates the choice set, it is the only one to verify the disjunctive decision rule, and the LogSum is low. Conversely, if K different alternatives perform best in the K attributes, the LogSum is at its highest (it is bounded between 0 and $\ln(K)$). However, this property may lead to counter-intuitive results if we assume the LogSum as a welfare metric. Fig. 9 plots the LogSum using the example from Table 5. We see that decreasing the cost and travel time of Alternative 3 may decrease the LogSum, as this Alternative becomes dominant in the choice set. In general, we observe a non-monotonicity of the LogSum with respect to the attribute values, similarly to what Chorus (2012a) for the RRM LogSum. Thus, we do not advise to use the GRDM LogSum as a welfare metric.

3.4 Elasticities

Elasticities measure the relative response of a variable (e.g., the choice probabilities) to the change of another (e.g., an attribute). For an alternative i of a choice set $\mathscr{C}$, we define the disaggregate direct point elasticity as the elasticity of the choice probability to the kth attribute of an alternative, it is calculated as:

$$E_{x_{ik}}^{P_i} = \frac{\partial P_i}{\partial x_{ik}}\frac{x_{ik}}{P_i} = \frac{\partial \ln P_i}{\partial x_{ik}} x_{ik} \tag{9}$$

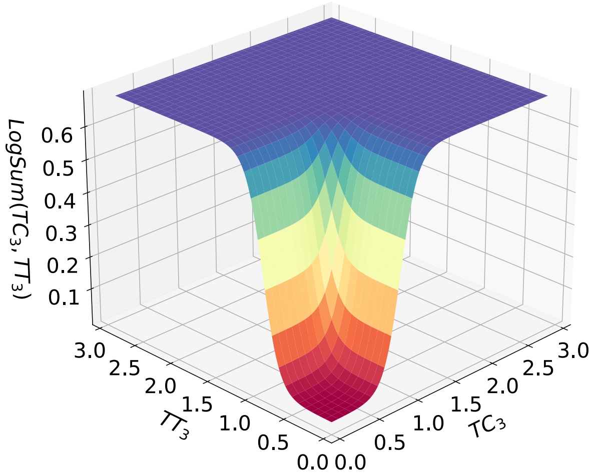

Fig. 9. LogSum given that $\alpha_{TT} = \alpha_{TC} = -10$, $\lambda_{TT} = \lambda_{TC} = 1$.

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

$$E^{P_i}_{x_{jk}} = \frac{\partial P_i}{\partial x_{jk}} \frac{x_{jk}}{P_i} \tag{10}$$

where $j \in \mathscr{C}$ is another alternative. The general formula for the direct and cross elasticities is given by (see Appendix C for a proof):

$$E^{P_i}_{x_{jk}} = \frac{\partial P_i}{\partial x_{jk}} \frac{x_{jk}}{P_i} = \lambda_l \alpha_l P_i^{-1} x_{jk} \sum_{q \in \mathscr{C}} (\delta_{iq} - P_i) \frac{P_{ql}}{1 - P_{ql}} (1 - P_q)(\delta_{qj} - P_{ql}) \tag{11}$$

where $P_{jl} = \exp(\alpha_l x_{jl}) / \sum_{q \in \mathscr{C}} \exp(\alpha_l x_{ql})$.

4 The need for multiple decision rules

This section demonstrates the need to consider multiple decision rules when capturing choice behaviour over a series of situations. Specifically, we demonstrate how compensatory choice models (RUMs) can struggle to capture non-compensatory choice behaviour and vice versa. Throughout the section, we demonstrate behaviour for the example in Table 1, where the third alternative is a compromise between the first two alternatives.

4.1 Compensatory choice models cannot always capture non-compensatory choice behaviour

We first show that a non-compensatory choice strategy can violate necessary conditions for a RUM, thereby showing that a RUM can struggle to capture non-compensatory behaviour. Block (1974) and Marschak (1959) give a set of testable properties of RUMs, which are summarised in Batley and Hess (2016). One of them is regularity: enlarging the set of feasible alternatives decreases the choice probabilities of the alternatives in the initial set. If $C_0 \subseteq \mathscr{C}$, let us define $p_{C_0}(i) = \mathbb{P}(U_n(\mathbf{x}_i) = \max_{j \in C_0} U_n(\mathbf{x}_j))$ for all $i \in C_0$. The regularity property can be defined as:

$$C_1 \subseteq C_0 \subseteq \mathscr{C} \implies \forall i \in C_1,\ p_{C_1}(i) \geq p_{C_0}(i) \tag{12}$$

Assume that an individual faces the choice situation presented in Table 1, and adopts the following non-compensatory choice strategy:

1. They may choose an alternative if it is the best in at least one attribute
2. If there is a tie, they will choose the alternative whose sum of attributes is the smallest
3. If, again, there is a tie, they will choose any alternative with equal probability

Assuming that individuals may not always be choosing from the full choice set $\mathscr{C}$, but that one alternative may be unavailable. Under this decision strategy, we can calculate the choice probabilities under the availability constraints (Table 6).

We observe, for instance, that $p_{\{1,2,3\}}(1) > p_{\{1,3\}}(1)$, and that $p_{\{1,2,3\}}(2) > p_{\{2,3\}}(2)$, which is inconsistent with the regularity property from Eq. (12), and thus with utility maximization. This means that making alternative 2 unavailable will make alternative 3 more competitive against alternative 1. This strategy, however, will still be consistent with the disjunctive decision rule.

Table 6
Observed shares for given choice sets, assuming the decision makers use a disjunctive strategy.

C	$p_C(1)$	$p_C(2)$	$p_C(3)$
$\{1,2,3\}$	0.5	0.5	0
$\{1,2\}$	0.5	0.5	–
$\{1,3\}$	0	–	1
$\{2,3\}$	–	0	1

Table 7
Estimation results with observed shares.

Attribute	MNL	RDM	GRDM ($\alpha \vert \lambda$)	
Travel Time	2.66	−91.75	−244.9	0.081
Travel Cost	3.02	−80.16	−183.5	0.031
Final LL	**−3012.5**	**−2079.4**		**−2059.3**
Adj. ρ^2	**0.087**	**0.370**		**0.376**

We next demonstrate how estimating a RUM on a dataset where decision-makers employ a non-compensatory strategy can lead to counter-intuitive results. Returning to the small choice example from Table 1, we demonstrate how estimating MNL on a dataset violating utility maximization will not converge to meaningful estimates. Imagine we estimate a linear MNL on a dataset where people adopt the non-compensatory decision strategy described above. The utilities can be written, for $i = \{1,2,3\}$, $U_i = V_i + \epsilon_i = \theta_{TT} TT_i + \theta_{TC} TC_i + \epsilon_i$. If we observe $N_1 > 0$ choices of 1 and $N_2 > 0$ choices of alternative 2, but no choice of alternative 3, $N_3 = 0$, we get the following log-likelihood function:

$$
\begin{aligned}
LL(\theta_{TT},\theta_{TC}) &= \sum_{n=1}^{N_1} \log(P_1(\theta_{TT},\theta_{TC})) + \sum_{n=1}^{N_2} \log(P_2(\theta_{TT},\theta_{TC})) \\
&= N_1 \log(P_1) + N_2 \log(P_2) \\
&= N_1\theta_{TC} + N_2\theta_{TT} - (N_1 + N_2)\log(e^{\theta_{TT}} + e^{\theta_{TC}} + e^{(0.2\theta_{TT}+0.2\theta_{TC})})
\end{aligned}
$$

Maximizing the likelihood function means setting the partial derivatives to 0, i.e.,

$$
\begin{aligned}
\frac{\partial LL}{\partial \theta_{TT}} = 0 &\iff N_2 - (N_1 + N_2)(P_2 + 0.2P_3) = 0 \\
\frac{\partial LL}{\partial \theta_{TC}} = 0 &\iff N_1 - (N_1 + N_2)(P_1 + 0.2P_3) = 0 \\
\text{Moreover, we have:} &\quad P_1 + P_2 + P_3 = 1
\end{aligned}
$$

These three equations give that $P_1 = N_1/(N_1 + N_2)$, $P_2 = N_2/(N_1 + N_2)$ and that $P_3 = 0$ (i.e., the likelihood estimator reproduces the observed shares). It follows that $V_1 - V_3$ and $V_2 - V_3$ must tend to $+\infty$, while V_1 and V_2 must keep values consistent with the observed shares, i.e., $V_1 - V_2 = \ln\left(N_1/N_2\right)$, which gives the following system of equations that θ_{TT} and θ_{TC} should verify:

$$
\left.
\begin{aligned}
-0.2\theta_{TT} + 0.8\theta_{TC} &= +\infty \\
0.8\theta_{TT} - 0.2\theta_{TC} &= +\infty \\
\theta_{TT} - \theta_{TC} &= \ln\left(\frac{N_1}{N_2}\right)
\end{aligned}
\right\}
$$

This system has no solution. When performing an estimation of the MNL on observed choices for which no one ever chooses the compromise alternative, the estimated parameters may lead to non-sensical estimates. Let us estimate a MNL with the observed choices from Table 3. The model estimates are given in Table 7. We observe that the MNL accommodates the absence of choice for the compromise alternative as a preference for more travel time and more travel cost. In contrast, it should be interpreted as a decision strategy for which individuals are averse to making compromises. Moreover, the MNL is outperformed by the non-compensatory RDM and GRDM, with its fit similar to the Null model.

While this example is hypothetical, this non-compensatory choice behaviour occurs in real-life situations. Some Stated Preference experiments (e.g., McCausland et al. (2020), Sælensminde (2006)) highlighted that, with high confidence, people were using decision strategies that violated utility maximization. Other examples of ways RUM assumptions are violated can be found in the literature (see, e.g. Hess et al. (2018) for a review). Both the RDM and GRDM may also violate the RUM regularity property from Eq. (12) for some values of parameters.

4.2 Non-compensatory models cannot always capture compensatory choice behaviour

Conversely to what was hypothesised in the previous subsection, we now test if the GRDM is consistent with choice behaviour if decision-makers mostly choose compromise alternatives (for instance, in the example from Table 1). Using the example from Table 1, the probability of alternative 3 has an upper bound that is strictly lower than 1. The probability of alternative 1 being

Table 8
Final likelihood when estimating MNL and GRDMs with different market share, using the example from Table 3 with different shares of observations.

N_1	N_2	N_3	LL_{MNL}	LL_{GRDM}
1700	1300	0	−3012.5	−2059.3
1200	900	900	−3271.5	−3258.9
300	300	2400	−2161.8	−2840.6

chosen for at least one attribute is higher than $1-(1-1/3)^{\lambda_1}(1-0)^{\lambda_2} = 1-(2/3)^{\lambda_1}$, because it has a probability of at least 1/3 of being chosen for its travel time, as it is the best performing from the choice set. For the same reason, the probability of alternative 2 being chosen for at least one attribute is higher than $1-(2/3)^{\lambda_2}$. The probability of alternative 3 being chosen for at least one attribute is lower than $1-(1-1/2)^{\lambda_1}(1-1/2)^{\lambda_2} = 1-(1/2)^{\lambda_1+\lambda_2}$, as it is not the best in travel time (and thus has a lower probability than alternative 1 to be chosen for this attribute) nor in travel cost. Consequently, each alternative probability respects the following inequalities:

$$
\begin{aligned}
P_1^{\text{GRDM}} &\geq \frac{1-(2/3)^{\lambda_1}}{3-(2/3)^{\lambda_1}+(2/3)^{\lambda_2}+(1/2)^{\lambda_1+\lambda_2}} \\
P_2^{\text{GRDM}} &\geq \frac{1-(2/3)^{\lambda_2}}{3-(2/3)^{\lambda_1}+(2/3)^{\lambda_2}+(1/2)^{\lambda_1+\lambda_2}} \\
P_3^{\text{GRDM}} &\leq \frac{1-(1/2)^{\lambda_1+\lambda_2}}{3-(2/3)^{\lambda_1}+(2/3)^{\lambda_2}+(1/2)^{\lambda_1+\lambda_2}}
\end{aligned} \tag{13}
$$

Maximizing Eq. (13) with respect to λ_1 and λ_2 gives that $P_3^{\text{GRDM}} \leq 0.6304$. This maximum is attained when λ_1 and λ_2 approach 0. Conversely, a linear MNL can allocate a probability to alternative 3 higher than $1-\delta$ for any $\delta \in]0,1[$ e.g., using a linear additive specification with high values of θ_{TT} and θ_{TC}. Thus, if decision-makers adopt compensatory behaviour and mainly (or only) choose the compromise alternative, MNL will have a better fit than the RDM or the GRDM. In a nutshell, compensatory choice models fail when explaining some non-compensatory choice strategies, and non-compensatory choice models such as the GRDM fail when explaining compensatory choice strategies. However, different decision-makers or choice situations may warrant using either of these two strategies in real-life choice situations. Thus, combining compensatory and non-compensatory decision rules in a single model may be a practical approach to capture both of them.

5 Combination of decision rules in a single model

In the examples presented throughout the last section (see Tables 1 and 5), we assumed that we observed choices where the compromise alternative was never or usually chosen. In practice, when observing a dataset, it is impossible to know if individuals' choices result from using different decision rules or heterogeneous preferences. An estimation under different market shares (see Table 8 for final log-likelihoods) on the example shows that the MNL and the GRDM better explain different observed shares, depending on how many times the compromise alternative was chosen.

If these two models seem to capture different types of choice behaviour, it is likely that combining the two decision rules into a single choice model will lead to an improved fit. In this section, we aim to combine the two decision rules into a choice model using a LCCM.

Latent class choice models (LCCMs): The LCCM was first developed by Kamakura and Russell (1989). They assume that discrete classes can model heterogeneity between individuals or observations, each class representing different tastes or decision rules. The class membership being unobserved, it is modelled as a latent variable, modelled probabilistically. This model can be seen as a Mixed Logit with a finite mixing distribution but can also accommodate various decision rules (as did Hess et al. (2012)). The LCCM probabilities, for an alternative $i \in \mathscr{C}$, is given by the sum over all classes of the joint probability of each model and class, i.e.,

$$\mathbb{P}(y_{in}=1) = \sum_{m=1}^{M} \mathbb{P}(y_{in}=1, m) = \sum_{m=1}^{M} \mathbb{P}(y_{in}=1 \mid m)\mathbb{P}(m)$$

If $P_i^m(\beta_m) := \mathbb{P}(y_{in}=1 \mid m)$ is the probability of alternative according to a model m with parameters β_m and $\boldsymbol{\pi} = (\pi_1 \cdots \pi_{M-1}) \in [0,1]^{M-1}$ are the class allocation probabilities, which the model must estimate, $P_i := \mathbb{P}(y_{in}=1)$, can be rewritten as:

$$P_i = \sum_{m=1}^{M} \pi_m P_i^m(\beta_m), \quad \sum_{m=1}^{M} \pi_m = 1$$

Likelihood: Let us assume we observe N individuals, where individual n has T_n observations. For an observation $t \in \{1,\ldots,T_n\}$, we note $j_{nt} \in \mathscr{C}_{nt}$ the index of the chosen alternative in the choice set $\mathscr{C}_{nt}$. The log-likelihood function can be written as:

$$LL_{LC}(\beta_1,\ldots,\beta_M) = \sum_{n=1}^{N}\sum_{t=1}^{T_n} \log\left(\sum_{m=1}^{M} \pi_m P_{j_{nt}}^m(\beta_m)\right), \quad \sum_{m=1}^{M} \pi_m = 1 \tag{14}$$

With this formulation, we do not assume that the same individual always uses the same decision rule (i.e., is always allocated to the same class). This model thus allows for intra-personal heterogeneity.

Estimation: Standard Maximum Likelihood Estimation (MLE) procedures can be used to estimate parameters of the LCCM, i.e. by adopting some optimisation algorithm that maximises the LCCM likelihood formulation in Eq. (14). For the experiments in this paper, we utilised the L-BFGS-B minimisation algorithm, where we minimise the log-likelihood. For accuracy and to speed up estimation times, we used the analytical gradient computation (see Appendix B). LCCMs are not convex models, meaning the traditional gradient-based MLE may converge to local optima (Train, 2008). To deal with this problem, we used several random starting values for the MLE algorithm and showed the model with the best likelihood, similarly as Hess et al. (2012). As the models contained only two classes, the estimates were relatively stable.

5.1 Posterior analysis

Calculating the class allocation posterior probabilities is possible using Bayes' theorem. This posterior reflects the probability of an observation being explained by one of the models. The posterior class membership probability $\mathbb{P}(m \mid y_{j_{nt}} = 1)$ of an observation $j_{nt} \in \mathscr{C}_{nt}$ is given by:

$$\mathbb{P}(m \mid y_{j_{nt}} = 1) = \frac{\mathbb{P}(y_{j_{nt}} = 1 \mid m)\mathbb{P}(m)}{\mathbb{P}(y_{j_{nt}} = 1)} = \frac{\mathbb{P}(y_{j_{nt}} = 1 \mid m)\mathbb{P}(m)}{\sum_{m'} \mathbb{P}(y_{j_{nt}} = 1 \mid m')\mathbb{P}(m')}$$

This means that for observation t of individual n, the class membership posterior probability for class m_0 can be calculated as:

$$\tilde{\pi}_{m_0}^{j_{nt}} = \frac{\hat{\pi}_{m_0} P_{j_{nt}}^{m_0}(\hat{\beta}_{m_0})}{\sum_{m=1}^{M} \hat{\pi}_m P_{j_{nt}}^{m}(\hat{\beta}_m)} \tag{15}$$

where $\hat{\boldsymbol{\pi}}, \hat{\beta}_1, \dots, \hat{\beta}_M$ are the LCCM estimated parameters.

We can derive the expected individual MRS for each observation $\mathbb{E}\left[\mathrm{MRS}_{kl}^{j_{nt}}\right]$ based on the posterior class membership probabilities:

$$\mathbb{E}\left[\mathrm{MRS}_{kl}^{j_{nt}}\right] = \sum_{m=1}^{M} \tilde{\pi}_{m_0}^{j_{nt}} \mathrm{MRS}_{kl,m}^{j_{nt}} \tag{16}$$

where $\mathrm{MRS}_{kl,m}^{j_{nt}}$ is the calculated marginal rate of substitution between two attributes k and l of the chosen alternative according to model m.

6 Case studies

In this section, we estimate models developed in this paper on two different route choice datasets. Route choice is a complex task to model, as, for example, transport networks allow for many different route alternatives between origin and destination. Travellers will typically compare a subset of these routes but may not have the time or knowledge to weigh all their attributes. Route choice has thus the potential to exhibit multiple decision rules across and among the users, which will be explored by the combined LCCMs estimations.

6.1 Case study 1: Bicycle route choice in the greater copenhagen area

Our first case study models cyclists' route choices in the Copenhagen Metropolitan area.

6.1.1 The data

We utilised a large-scale crowdsourced dataset of bicycle GPS trajectories received from Hōvding.[5] The original dataset covers the entire Greater Copenhagen Area (see Fig. 10) 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 the supplementary information in Fosgerau et al. (2023). The final dataset for model estimation consists of a subset of this dataset containing 4355 trips made by 4355 cyclists.

The disaggregated network consists of 420,973 directed links and 324,492 nodes and relies on open-source network data from Open Street Map (OSM[6]). The OSM network is a free editable map of the world based on crowdsourced data and manual edits from millions of private users. It contains attributes about road surface (asphalt, gravel, or cobblestones), road importance (e.g., large road, residential road...), type of bicycle infrastructure (e.g., bike lane, segregated bike path).

5 https://hovding.com
6 www.openstreetmap.org

Fig. 10. Heatmaps of anonymised GPS trajectories from Hōvding.

Choice set generation: The method used for the choice set generation uses a novel approach. First, it uses the stochastic approach (Nielsen, 2004; Bovy and Fiorenzo-Catalano, 2007), drawing a large number of routes on the network between each OD. These routes are then filtered with a local optimality criterion (Abraham et al., 2013), defined as the minimum length of a subpath that is not the shortest path. This criterion constrains the smoothness of the routes and their mutual overlap. Appendix D gives a more detailed description of the choice set generation method.

6.1.2 Model specification

The attributes of an alternative i of a choice set $\mathscr{C}$ can be described as:

- L_i (km): total route length
- E_i (m): route elevation gain when steepness > 3.5%
- No_i (km): length of route segments where no bicycle infrastructure available
- S_i (km): route length on a non-asphalt surface (i.e. gravel, cobblestones)
- W_i (km): route length using wrong ways (cycling against traffic).

Path-Size correction: The Independence of Irrelevant Alternatives (IIA) property of RUMs is usually violated in route choice, as different routes from a choice set are often correlated through sharing links. Therefore, the route alternatives are not independent. It is well-known that the MNL RUM cannot account for route correlation, and numerous route choice models have been developed to address this; see, e.g., Duncan et al. (2020) for a recent review of such models. Due to the computational complexities in capturing route correlation on detailed networks (such as in this case study), a common approach has been to include heuristic path size correction factors within the deterministic utilities, i.e. with the Path-Size Logit model (Ben-Akiva and Bierlaire, 1999).

The Path-Size term for an alternative $i \in \mathscr{C}$ is defined as:

$$\gamma_i = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in \mathscr{C}} \delta_{aj}},$$

where Γ_i is the set of links for alternative i, l_a is the length of link a, L_i is the length of alternative i, and δ_{aj} equals 1 if j includes link a and 0 otherwise. This correction term penalises the utility of alternatives that overlap a lot, as they may not be considered different by the decision maker. The GRDM model does not have the same IIA property as the MNL, and it has yet to be explored what the correlation between alternatives means in the context of disjunctive models, as well as non-compensatory models in general, so we leave this to be investigated in future research.

We estimated the following models (all combined models are combined with a two-class LCCM at the observation level):

1. A simple MNL, with the following deterministic utility specification, for $i \in \mathscr{C}$:

$$V_i = \beta_L \times L_i + \beta_E \times E_i + \beta_{\text{No}} \times \text{No}_i + \beta_S \times S_i + \beta_W \times W_i + \beta_{\text{PS}} \ln(\gamma_i) \quad (17)$$

2. Combination of two MNL, called **MNL + MNL**, using the same specification (Eq. (17)) for both MNL classes.
3. A simple GRDM, including every attribute

Table 9
Model results (first part). The attributes are *Length*: route length; *Elev. Up*: vertical elevation gain when steepness is over 3.5%; *No*: route length without bicycle lane or path; *Surf. Other*: route length on a non-asphalt surface; *Wrong Way*: route length on wrong ways. We removed the insignificant parameters, which are marked with a dash.

Attribute	MNL	MNL + DDM	MNL + RDM		GRDM		MNL + GRDM		
Model	–	–	*MNL*	*RDM*	α	λ	MNL	*GRDM* $(\alpha\|\lambda)$	
Length (km)	−29.3	−34.41	−41.6	−405	−86.75	0.0032	−37.09	−238.09	0.280
Elev. Up (m)	−0.062	−0.0860	−0.132	−0.177	–	–	−0.267	–	–
Surf. Other (km)	−4.59	−4.546	−7.02	−22.1	−31.93	0.0443	−7.506	−20.36	0.046
BikeNo (km)	−4.65	−5.411	−6.58	−12.0	−11.55	0.0212	−6.730	−7.971	0.152
Wrong Way (km)	−8.47	−10.25	−13.4	−10.2	−32.7	0.04	−14.25	−6.958	0.116
ln(PS)	1.09	1.208	1.39				1.322		
π_m (first model)		0.861	0.807				0.726		
Sample size	4,355	4,355	4,355		4,355		4,355		
Nb. of parameters	6	7	12		8		15		
Final LL	−12,666	−11,973	−11,721		−14,127		−11,514		
Adj. ρ^2	0.471	0.500	0.510		0.410		0.519		
BIC	25,382	23,971	23,543		28,281		23,083		

4. Combination of an MNL and a deterministic disjunctive model, called **MNL + DDM**, using the specification from Eq. (17) for the MNL part.
5. Combination of a MNL and a RDM, called **MNL + RDM**, using the specification from Eq. (17) for the MNL part
6. Combination of a MNL and the GRDM, called **MNL + GRDM**, using the specification from Eq. (17) for the MNL part
7. Combination of two GRDM, called **GRDM + GRDM**

The combined MNL + MNL model allows for benchmarking the newly developed model and comparing models with the same number of classes.

6.1.3 Results

First, before we estimated the models on the dataset, simulation estimation experiments were conducted, similarly to Duncan et al. (2020), Duncan et al. (2022). Their purpose was to assess whether the GRDM could be successfully estimated and, specifically, to investigate whether there may be any identification issues between the inverse standard deviation of the error terms ($\boldsymbol{\alpha}$) and the exponents (λ), as both may have similar effects on choice probabilities. To test this, we conducted the following steps. We first assumed true model parameters, then drew observations on choice situations from this case study according to the GRDM probabilities and the assumed true parameters. We then re-estimated the GRDM on this generated choice data. The simulation estimation experiments were repeated 100 times, where the results showed no identification problems, i.e., the model could replicate the true values for $\boldsymbol{\alpha}$ and λ.

We now proceed to display results from the estimation work on the dataset. To evaluate the performance of the models, the following goodness-of-fit metrics have been calculated: the adjusted McFadden rho-squared (McFadden, 1974a) and the Bayesian Information Criterion (BIC, Schwarz (1978)), given by

$$\text{Adj. } \rho^2 = 1 - \frac{LL(\hat{\beta}) - k}{LL(0)}$$
$$\text{BIC} = k\ln(N) - 2LL(\hat{\beta})$$

where N is the number of observations in the model, k is the number of estimated parameters, $LL(\hat{\beta})$ is the model final log-likelihood, and $LL(0)$ is the likelihood of the null model. Additionally, significance t-tests were conducted for each estimated parameter. For the MNL parameters and the GRDM exponents (i.e., λ), the estimated parameters were tested against 0. For the GRDM variances (i.e., $\boldsymbol{\alpha}$), the estimated parameters were also tested against ∞, i.e. $\forall k,\ 1/\alpha_k$ was tested against 0. These tests were computed by calculating the Hessian matrix of the likelihood function around the value of the estimates, performing a numerical differentiation of the analytical gradients (see Appendix B for the calculation of the analytical gradient). The models were first estimated using the full specification. Then, insignificant parameters were removed, and the models were re-estimated. The results are presented in Tables 9 and 10.

Each model is nested into one another (the RDM collapses to the DDM by setting all its parameters to $+\infty$, and the GRDM collapses to the RDM by setting all its exponents to 1), so we can perform Likelihood Ratio Tests (LRTs) on any pair of models:

$$\text{MNL + MNL vs. MNL} : \mathbb{P}\left[\chi^2(6) < -2(LL_{MNL} - LL_{MNL+MNL})\right] = 0$$
$$\text{MNL + DDM vs. MNL} : \mathbb{P}\left[\chi^2(1) < -2(LL_{MNL} - LL_{MNL+GRDM})\right] = 0$$
$$\text{MNL + RDM vs. MNL + DDM} : \mathbb{P}\left[\chi^2(5) < -2(LL_{MNL+DDM} - LL_{MNL+RDM})\right] = 5.43 \times 10^{-107}$$
$$\text{MNL + RDM vs. MNL + GRDM} : \mathbb{P}\left[\chi^2(5) < -2(LL_{MNL+GRDM} - LL_{MNL+RDM})\right] = 1.42 \times 10^{-87}$$

The LRTs show that including multiple classes can improve the model fit significantly. Furthermore, adding random error components to the perceived attribute value (RDM vs. DDM) and exponents (GRDM vs. RDM) further improves the fit. The Adjusted

Table 10
Model results (part 2), after removing the insignificant parameters, which are marked with a dash.

Attribute	MNL + MNL		GRDM + GRDM			
Model	*MNL 1*	*MNL 2*	*GRDM 1* $(\alpha\|\lambda)$		*GRDM 2* $(\alpha\|\lambda)$	
Length (km)	−67.04	−9.497	−233.0	0.0569	−56.16	10.64
Elev. Up (m)	−0.620	–	–	–	–	–
Surf. Other (km)	−9.353	−2.026	−27.88	0.00531	–	–
BikeNo (km)	−5.750	−3.942	−11.75	0.0911	–	–
Wrong Way (km)	−13.43	−5.282	−19.63	0.0342	−9.22	0.00898
ln(PS)	1.873	0.806				
π_m (first model)	0.677		0.765			
Sample size	4355		4,355			
Nb. of parameters	12		13			
Final LL	−11,903		−13,907			
Adj. ρ^2	0.502		0.419			
BIC	23,915		27,861			

Table 11
Average log-likelihood on the cross-validation sets.

Model	MNL	GRDM	MNL + MNL	MNL + GRDM	GRDM + GRDM
Average LL	−6,359.11	−7109.28	−6001.14	−5,769.20	−7004.43

rho-squared and BIC metrics confirm this trend. When estimating a Latent Class model with each class using a different decision rule, Hancock and Hess (2021) highlighted the risk of these models uncovering taste heterogeneity rather than heterogeneity in decision rule, and showed evidence of this pattern in RUM-RRM mixtures. However, the better fit of the MNL + GRDM than the MNL + MNL suggests that this model captures not only taste heterogeneity but heterogeneity in the decision rule.

Even when using a disjunctive rule, cyclists, in aggregate, tend to favour some attributes over others. It is an essential finding that while the GRDM alone and two-class perform poorly, its combination with MNL is the best fit for the data. This provides empirical support for the hypothesis that behaviour over a series of choice situations may be best captured by considering that individuals might adopt different decision rules in different choice situations, such as compensatory or non-compensatory.

We see from Tables 9 and 10 that the model estimates for the class allocation probabilities (π_{MNL}) give a higher share for the MNL than for the non-compensatory counterpart (86.1% for the MNL + DDM, 80.7% for the MNL + RDM, and 72.6% for the MNL + GRDM). Hence, most observed choices are better explained by utility maximization than by a disjunctive decision rule. However, adding flexibility to the disjunctive models, using random error terms and exponents for attributes, allows the non-compensatory part to explain a larger share of cyclists' choices.

The significant parameters' signs and magnitudes all make sense. For instance, MNL will predict a dispreference for routes without bicycle infrastructure and are willing to detour around 15.9% to avoid them according to the MNL, or 18.1% according to the MNL part of the MNL + GRDM model. In general, in combined models, the MRS for Length of the MNL component are higher than in non-combined models, meaning that the GRDM captures some of the observations that strongly prefer Length and are insensitive to the other attributes. In contrast, the MNL part captures more individuals with a stronger preference for short distances.

For the GRDM component of the MNL + GRDM model, the $\boldsymbol{\alpha}$ parameter shows a lower variance for the length error term than for the other attributes. This indicates a minor variability in how people perceive the shortest route as the best performing one compared, for instance, to the amount of unprotected bicycle infrastructure. It makes sense from a behavioural point of view, as decision-makers probably have better knowledge on the total route distance than on other attributes. Regarding attribute relative importance (the λ parameter), we see that length is the most decisive attribute, followed by length on bicycle infrastructure, wrong ways, and on a non-asphalt surface. Elevation gain does not influence choice behaviour for the GRDM part.

6.1.4 Model validation

In this subsection, we assess the generalizability of the different estimated models to ensure these do not overfit the data. We tested each model's forecasting ability by performing a Monte Carlo cross-validation. To do so, we repeated $N = 10$ times the following steps:

1. Randomly split the original dataset S into a training S_t and validation set S_v ($|S_t| = |S_v| = 0.5|S|$).
2. Estimate all the models on the training set S_t, obtain, for each model m, the training parameters β^t_m.
3. Calculate, for each model m, the log-likelihood on the validation set, $\text{LL} = \sum_{x\in S_v} \log P^m_{i_x}(\beta^t_m)$

where i_x is the chosen alternative index for observation x, and $P^m_{i_x}(\beta_m)$ is its choice probability according to the model m and its estimated parameters β^t_m. Fig. 11 and Table 11 show the cross-validation results. First, we see that the predictive ability of the models is similar to what indicated the model fit results. Accounting for heterogeneous decision rules allows for a better model generalization than not doing so and a better one than only accounting for taste heterogeneity. The MNL + GRDM was the best-performing model in all experiments, suggesting that this decision rule combination is the most promising.

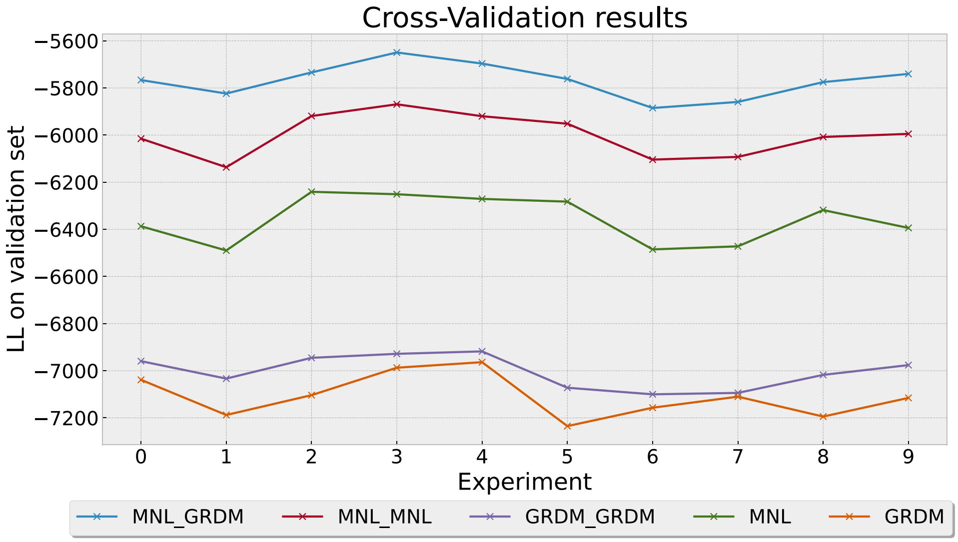

Fig. 11. Cross-validation results.

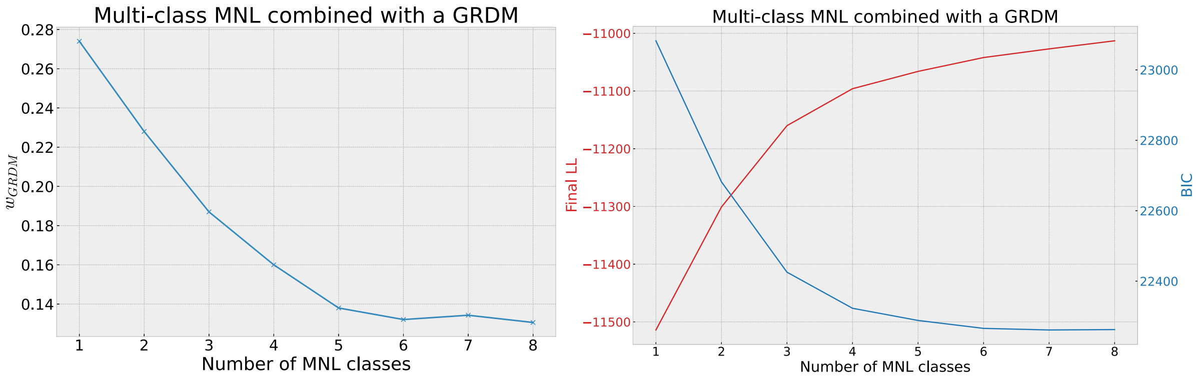

Fig. 12. Left: Weight of the GRDM class as the number of MNL classes increases. Right: Final log-likelihood and BIC in function of the number of MNL classes.

6.1.5 Decision rule heterogeneity, taste heterogeneity, or non-linear preferences?

While the MNL + GRDM outperforms the two-class MNL in terms of fit and prediction ability, it is hard to know if these improvements come from the capture of different tastes or non-linear preferences by the GRDM. One hypothesis could be that extreme tastes or non-linear sensitivities not captured well by the MNL class could have been better captured by the GRDM class. This subsection tests this hypothesis by estimating several additional models, including taste heterogeneity or non-linear sensitivities.

Decision rule or taste heterogeneity?

First, we try to disentangle between decision rule and taste heterogeneity by estimating models that account for both, i.e., we combine several MNL classes and one GRDM in a Latent Class framework. This allows testing the impact of adding more MNL classes on the GRDM's weight in the Latent Class model. If this is the case, it would mean that the GRDM may have captured some compensatory behaviour with different tastes. We kept increasing the number of MNL classes until the model BIC stopped decreasing and added up to eight MNL classes, so nine classes in total. It is important to note that adding more classes in LCCMs potentially increases the number of local optima for the log-likelihood function. The more classes, the more times we had to repeat the estimations with random initial conditions. Fig. 12 summarises the evolution of the model fit and the GRDM weight as the number of classes increases. We can observe that adding more MNL classes decreases the GRDM weight. This suggests that the GRDM possibly captured part of what can also be described as taste heterogeneity. We also observe, though, that the GRDM weight decrease slows down when adding more than five MNL classes, which suggests that there is a significant amount of observations (around 13%) that a utility-maximization decision rule cannot explain, while a disjunctive decision rule can. Moreover, we noticed that some MNL classes also showed a few positive coefficients, i.e., sensitivities to attributes like elevation gain or absence of bicycle infrastructure. These parameters, which are expected to be negative, can also allow MNL to accommodate non-compensatory behaviour, as we highlighted in Section 4.

Table 12
Model estimates for the models which transformed the "Length" attribute with a Logarithm or Box–Cox function in the MNL component.

	MNL (log Length)	MNL (Box–Cox Length)	MNL (Box–Cox Length) + GRDM		
Model			*MNL*	*GRDM* $(\alpha\|\lambda)$	
Length (km)	−49.96	−43.30	−46.63	−249.89	0.3635
Elev. Up (m)	−0.0704	−0.0562	−0.1189		
Surf. Other (km)	−4.332	−4.436	−7.133	−51.33	0.0493
BikeNo (km)	−4.497	−4.496	−6.446	−8.500	0.1643
Wrong Way (km)	−8.135	−8.210	−13.024	−6.394	0.1492
ln(PS)	0.7479	1.045	1.215		
Box–Cox parameter μ	–	0.4313	0.5509		
π_m			0.756		
Nb of parameters	6	7	16		
Final LL	−12,473	−12,334	−11,452		
Adj. ρ^2	0.4790	0.4847	0.5212		
BIC	24 967.83	24 693.47	22962.22		

Decision rule heterogeneity or non-linear preferences?

We estimated models incorporating a logarithmic, or Box–Cox transformed "Length" attribute for the MNL component to disentangle between decision rule heterogeneity. We tested the influence of this inclusion on the fit and weight of the GRDM in a combined model. The Box–Cox transformation (Box and Cox, 1964) is defined as:

$$\text{Box–Cox}(x,\mu) = \begin{cases} \frac{x^{\mu}-1}{\mu} \text{ if } \mu \neq 0 \\ \ln(x) \text{ if } \mu = 0 \end{cases}$$

For the logarithmic model, the MNL deterministic utility is thus given by:

$$V_i = \beta_L \times \ln(L_i) + \beta_E \times E_i + \beta_{\text{No}} \times \text{No}_i + \beta_S \times S_i + \beta_W \times W_i + \beta_{\text{PS}} \ln(\gamma_i)$$

For the Box–Cox models, the MNL component deterministic utility is thus given by:

$$V_i = \beta_L \times \text{Box–Cox}(L_i,\mu) + \beta_E \times E_i + \beta_{\text{No}} \times \text{No}_i + \beta_S \times S_i + \beta_W \times W_i + \beta_{\text{PS}} \ln(\gamma_i)$$

where μ is a parameter to estimate. Table 12 shows the estimation results from the non-linearly specified models. We compare these results from the baseline models in Table 9. The MNL models using a logarithmic or Box–Cox specification for the "Length" attribute outperformed the ones that assumed a linear sensitivity. However, the improvement in fit is much lower than the one of adding a GRDM component, which implies that the GRDM captures more than this non-linearity. Moreover, combining a GRDM with a Box–Cox-specified MNL drastically improves the model fit compared to a single Box–Cox-specified MNL. The GRDM weight is similar to the MNL + GRDM from Table 9. These observations imply that the GRDM does not pick up non-linear preferences, but rather decision rule heterogeneity.

6.1.6 Posterior analysis

We can calculate the posterior class membership probabilities of the MNL + GRDM combination to explore the choice situations in which the GRDM or the MNL explain the choices best.

Willingness To Detour (WTD): We define a cyclist's willingness to detour for another attribute as the MRS of an attribute with length. If an attribute's WTD is positive, it corresponds to the relative amount of length one is willing to cycle to avoid this attribute (e.g., a non-smooth road surface). If it is negative, it corresponds to the relative length one is willing to cycle to get this attribute (e.g., a segregated bicycle path). We can calculate the a posteriori expected WTD for each observation based on the posterior probability of class membership (see Eqs. (15) and (16)).

The distributions from Fig. 13 show heterogeneity in the expected value of the distance of every attribute for every observation. The median values of these WTDs are slightly higher than the ones of the initial MNL model estimates. We can note that, for some observations, cyclists have a marginal rate of substitution close to 0, meaning that they are unwilling to trade any cycling distance for another attribute (or are insensible to it). For some other observations, cyclists are willing to trade more distance for another attribute. This heterogeneity may come from tastes or the use of a different decision rule, meaning they are unwilling to trade attributes. Some observations cut off from the plots from Fig. 13, also have incredibly high WTD values. An extremely high WTD will likely be calculated in a choice situation in which the individual chooses an alternative for another attribute other than distance. According to the GRDM model, the decision-maker will not trade any amount of this attribute for a reduced distance, highlighting a non-compensatory behaviour.

The example from Table 13 shows a chosen route with a particularly high WTD for "Surface other than asphalt", indicating that the individual will not trade any added non-asphalt surface for a reduced length. The utility-maximizing alternative dominates the alternative, which is also the shortest. First, this alternative belongs to the GRDM class with a probability close to 1. While the MNL cannot explain this choice, the GRDM explains the choice because the decision-maker was only interested in minimizing his distance cycled on a non-asphalt surface, being insensitive to increases in length.

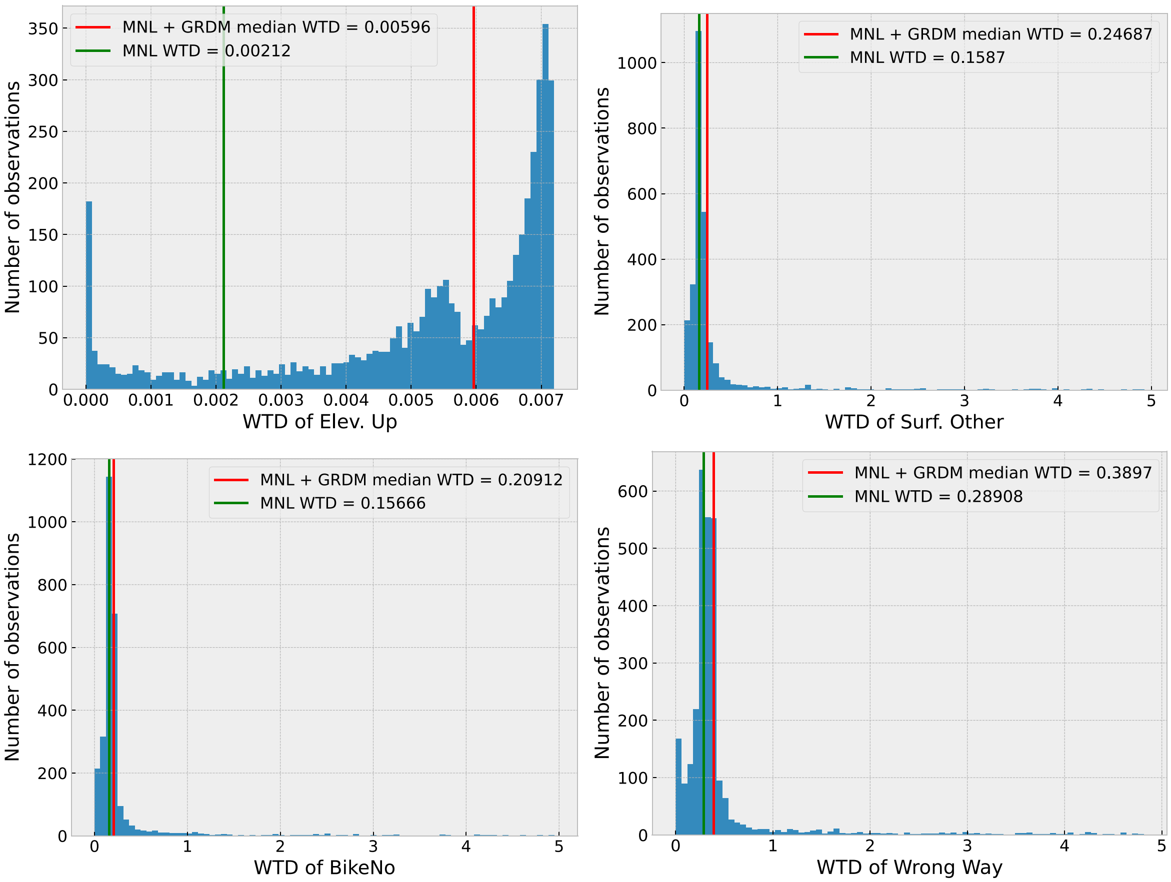

Fig. 13. Value of Distance distribution of the different attributes.

Table 13
Example of chosen route vs utility-maximizing route, very high WTD for riding on asphalt.

	Length	Elev. Up	Surf. Other	BikeNo	Wrong Way	ln(PS)
Chosen route	8.53	4.05	0	3.33	0.596	−3.04
Utility-maximizing route	8.27	4.05	0	0.78	0.12	−3.74

$\tilde{\pi}_{MNL} = 7.7e-12$; WTD of Surf. Other: 1.27e12.

Table 14
Example of chosen route vs utility-maximizing route, very high WTD for non-asphalt surface.

	Length	Elev. Up	Surf. Other	BikeNo	Wrong Way	ln(PS)
Chosen route	2.81	2.083	0	1.76	1.00	−6.72
Utility-maximizing route	2.90	1.88	0	0.86	0.071	−6.59

$\tilde{\pi}_{MNL} = 4.1e-8$; WTD of Elev. Up = 3.06e−10, WTD of Surf. Other = 4.4e−8, WTD of BikeNo = 2.25e−5, WTD of Wrong Way = 2.06e−7.

Conversely, in the example from Table 14, the WTD of every attribute is very low. This happens because the decision-maker chose the shortest route, while it is far from being the utility-maximizing route. We can interpret that the decision-maker only wanted to minimise length, regardless of the other attribute, and is thus not willing to make a detour for any quantity of another attribute.

6.1.7 Choice probabilities illustration

Fig. 14 shows a plot of the choice probabilities for the estimated MNL, GRDM, and combined MNL + GRDM with the estimated parameters from Table 9, using one of the observed ODs from the dataset. From these plots, we can see that the MNL choice probability increases directly as its length decreases. For the GRDM, however, this choice probability increases only when the alternative length becomes close to being the shortest in the choice set. The GRDM assumes decision-makers are almost insensitive to an attribute change until it becomes the best in the choice set, with the coefficient α influencing the slope of this increase.

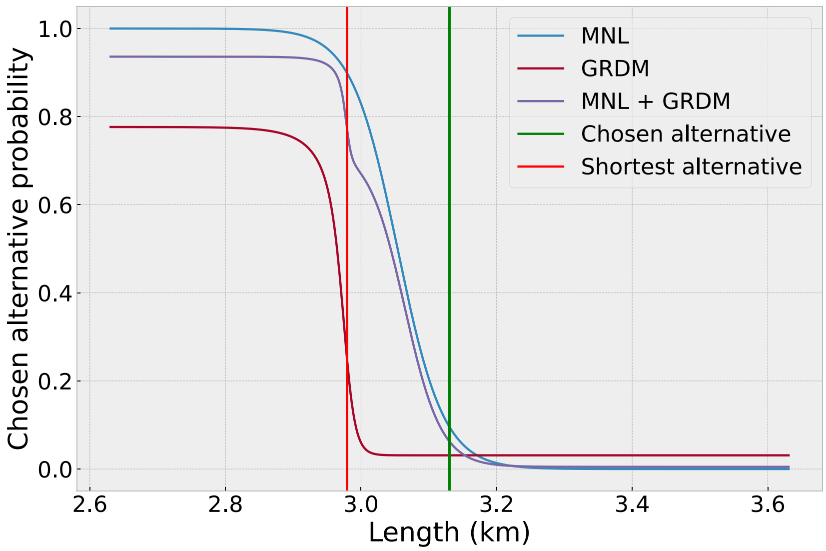

Fig. 14. Choice probability of the chosen alternative as a function of its length (the green line gives original length). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Another observation is that, whatever its length, the GRDM choice probability of the chosen alternative will neither tend to be zero nor one. This is linked to other alternatives performing best in other attributes, so the disjunctive model will allocate them non-zero probabilities. The choice probabilities do not tend to zero when Length increases because the chosen alternative also performs best in another attribute. This is a major difference with the MNL, for which the choice probability tends to zero with a smaller length and to one when the route length tends to infinity.

For the Latent Class MNL + GRDM, we observe a combination of the MNL and GRDM patterns. There are firstly two probability drops with increasing length. The first happens when the chosen route is no longer the shortest and is thus less favoured by the disjunctive decision rule. The second happens as the route deterministic cost increases, i.e., the route is less favoured by cost minimization. The choice probabilities limits lie between the MNL and GRDM ones.

6.2 Case study 2: Public transport route choice in the Greater Copenhagen area

In this second case study, the newly developed models are tested on a route choice model for the Greater Copenhagen Region's large-scale multimodal public transport network. The dataset includes metro, urban rail, local trains, regional trains and busses. 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 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 alternatives to the chosen route were generated using a Doubly-Stochastic method (Nielsen, 2004). A first model was estimated for the combined MNL + GRDM models, including all attributes for both models. All the insignificant parameters were removed, and a new reduced model was estimated.

6.2.1 Results

The model results are in Table 15.[7] The time attributes are expressed in minutes. All estimated parameters for the MNL part have the expected signs and relative magnitudes (see Nielsen et al. (2021)[8] for a further discussion of the MNL estimates). Combining the MNL and the GRDM significantly improves fit for both trip purposes. The π_m estimates indicate that even more than for cyclist's route choice, public transport user choices are better explained by the MNL component than by the GRDM component. This is more true for work trips (93.8%) than for leisure trips (88.9%). Similarly to the previous case study, we performed LRTs to compare the MNL + GRDM and the MNL:

$$\begin{aligned} \text{MNL + GRDM vs. MNL, Work} &: \mathbb{P}\left[\chi^2(13) < -2(LL_{MNL} - LL_{MNL+GRDM})\right] = 3.38 \times 10^{-217} \\ \text{MNL + GRDM vs. MNL, Leisure} &: \mathbb{P}\left[\chi^2(13) < -2(LL_{MNL} - LL_{MNL+GRDM})\right] = 2.55 \times 10^{-194} \end{aligned}$$

[7] The GRDM alone and the combine GRDM + GRDM were also estimated on this dataset but gave a much inferior fit to the data, so it is not presented here. Their performance on the cross-validation dataset is however presented in the next subsection.

[8] The original model omitted Alternative Specific Constants (ASCs) in its specification. Our MNL specification with ASCs improved the model fit and estimates plausibility compared to the cited work.

Table 15
Model results.

Model	MNL - Work	MNL - Lei.	MNL + GRDM - Work			MNL + GRDM - Leisure			MNL + MNL - Work		MNL + MNL - Leisure	
Attributes			*MNL*	*GRDM* $(\alpha\|\lambda)$		*MNL*	*GRDM* $(\alpha\|\lambda)$		MNL 1	MNL 2	MNL 1	MNL 2
Alternative Specific Constants												
Bus	−0.5147	−0.3498	−0.8993			−0.1145			−0.7662	−0.0389	−0.1063	−1.1002
Local train	−0.2486	0.9933	−3.3548			−0.4492			−0.6233	−0.5893	−0.6063	0.8681
Metro	2.0474	2.0759	2.4728			3.0081			2.3623	2.3094	3.1960	0.9203
Reg. and intercity train	−0.3639	−0.7410	−0.9470			−1.3671			−0.6272	0.0496	−3.316	0.6307
S-train	1.5846	1.5689	1.7331			1.8636			1.8097	1.1471	3.0864	0.2256
In-vehicle time												
Bus	−0.2486	−0.2005	−0.3451	–	–	−0.3162	–	–	−0.3251	−0.0882	−0.3672	−0.1014
Local train	−0.2370	−0.2110	−0.1918	–	–	−0.1270	–	–	−0.2531	−0.1305	−0.1273	−0.1544
Metro	−0.2854	−0.2395	−0.4146	–	–	−0.3840	–	–	−0.3452	−0.2403	−0.2832	−0.2250
Reg. and intercity train	−0.2251	−0.1912	−0.2630	−13.973	5.212e−3	−0.2200	−10.913	2.773e−3	−0.2445	−0.1358	−0.1650	−0.2395
S-train	−0.2649	−0.2129	−0.3476	–	–	−0.2750	–	–	−0.3152	−0.1491	−0.2930	−0.1902
Transfer components												
Transfer penalty	−2.5323	−2.5345	−3.0451	−14.080	3.601e−2	−3.452	−16.221	2.192e−2	−3.430	−0.6914	−4.635	−1.207
Transfer walking time	−0.1786	−0.2260	−0.2474	–	–	−0.3404	−13.751	5.057e−3	−0.2563	−0.1031	−0.4238	−0.0871
Transfer waiting time	−0.0469	−0.0431	−0.0581	–	–	−0.0605	–	–	−0.0536	−0.0428	−0.0613	−0.0451
Other components												
Access time	−0.4569	−0.4449	−0.6573	−35.124	2.162e−2	−0.6416	−17.933	1.418e−2	−0.6819	−0.0490	−0.8882	−0.1671
Egress time	−0.4002	−0.3829	−0.5628	−16.472	2.571e−2	−0.5920	−16.843	1.275e−2	−0.5936	−0.04289	−0.7769	−0.1378
Half of highest headway	−0.1006	−0.0948	−0.1733	−10.388	2.441e−3	−0.1517	−10.02	8.275e−4	−0.1728	0.0221	−0.1636	−0.0458
π_m			0.9380			0.8889			0.9545		0.8164	
Number of observations	2553	2257	2553			2257			2553		2257	
Number of parameters	16	16	27			24			23		23	
Null LL	−12,522	−10,770	−12,522			−10,770			−12,522		−10,770	
Final LL	−3,007	−3,346	−2,544			−2,957			−2,620		−3,032	
Adjusted rho-square	0.759	0.688	0.795			0.719			0.788		0.715	
BIC	6,148	6,826	5,314			6,148			5,517		6,341	

Table 16
Average log-likelihood on the cross-validation sets.

Model	MNL	GRDM	MNL + MNL	MNL + GRDM	GRDM + GRDM
Average LL (work)	−1,519.95	−3,482.64	−1,377.53	−1,310.62	−3,380.53
Average LL (leisure)	−1,693.79	−2,882.11	−1,583.56	−1,534.56	−2,777.38

These LRTs favour the model combination with high significance. We also see that combining two MNLs improves the fit to a lower degree than combining the MNL and the GRDM. In the following paragraphs, we interpret the estimates of the GRDM part for the best-performing models for Work and Leisure trips, i.e., the combination of MNL and GRDM.

MNL + GRDM - Work: First, for the error terms' variances (the $\boldsymbol{\alpha}$ parameter), we see the lowest variances for access and egress time. It could be explained by the better reliability of access and egress times or by a lower variability in people's dispreference for access and egress times. For the attributes' relative importance (the λ parameter), we observe that the Regional and Intercity Trains are the only transport modes for which in-vehicle time influences choice probabilities, which may be linked to the possibility that these types of modes are less punctual than other modes. The most important attributes, however, are access, egress time, and the presence of a transfer.

MNL + GRDM - Leisure: The error terms' variances are similar to the work trips. For the attributes' relative importance (the λ parameter), we observe that the Bus is the only transport mode for which in-vehicle time influences choice probabilities, which can be explained by a stronger aversion to travelling in buses than other modes.

6.2.2 Model validation

Similarly to the bicycle route choice case study, we perform the same cross-validation technique for the two public transport route choice datasets. The results are reported on Figs. 15 and 16 as well as on Table 16. We can see that the GRDM and the GRDM + GRDM perform poorly on this dataset (particularly for work trips). However, when combined with a MNL, this model becomes the best-performing amongst benchmarked models for every experiment when it comes to forecasting Public Transport's chosen routes.

6.2.3 Posterior analysis: relative values of time

Similarly to what was done for the bicycle route choice case study, we can infer the posterior MRS for each attribute. In this case, we can compare the relative values of time to a base reference (e.g., the access time). A relative VoT is the amount of time the decision-maker would be willing to spend for a reduced unit of time of the baseline. For instance, if the metro had a relative VoT of 0.5 to access time, it would mean that the decision-maker would trade two minutes more metro time for each minute of walk to the metro station. Table 17 shows each model's relative Values of Time to access time. While the MNL models give the same relative VoT for each observation, the LCCM gives a different value for each observation. Fig. 17 presents an example of the posterior distribution of the relative values of access time to egress time. It shows that the median value is slightly below one for both purposes, meaning individuals are willing to walk slightly longer from the station than to the station. The table thus reports the median of these values. The median is used instead of the mean as the GRDM relative Values of Time can approximate infinity.

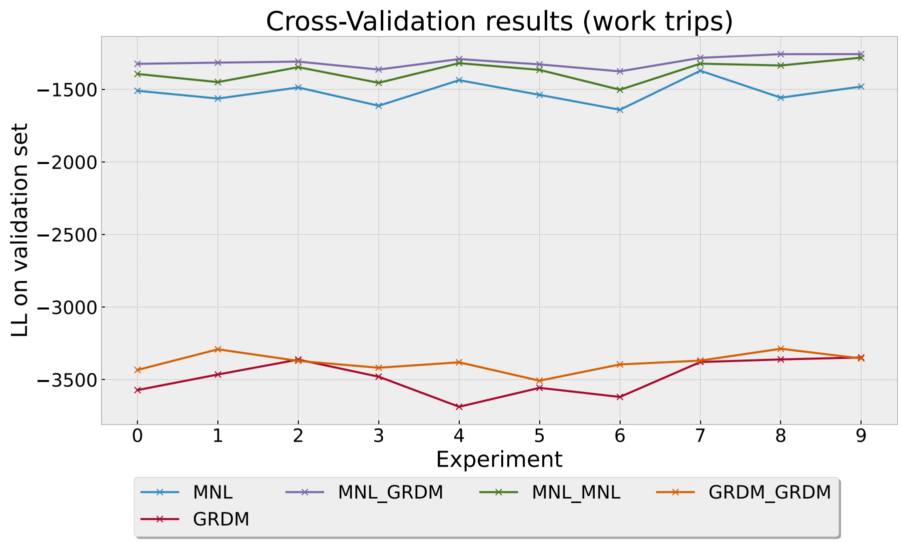

Fig. 15. Cross-validation results — Work trips.

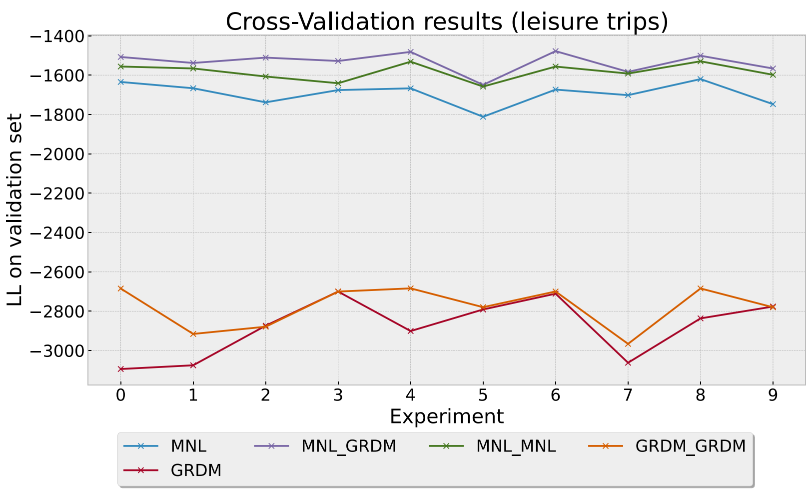

Fig. 16. Cross-validation results — Leisure trips.

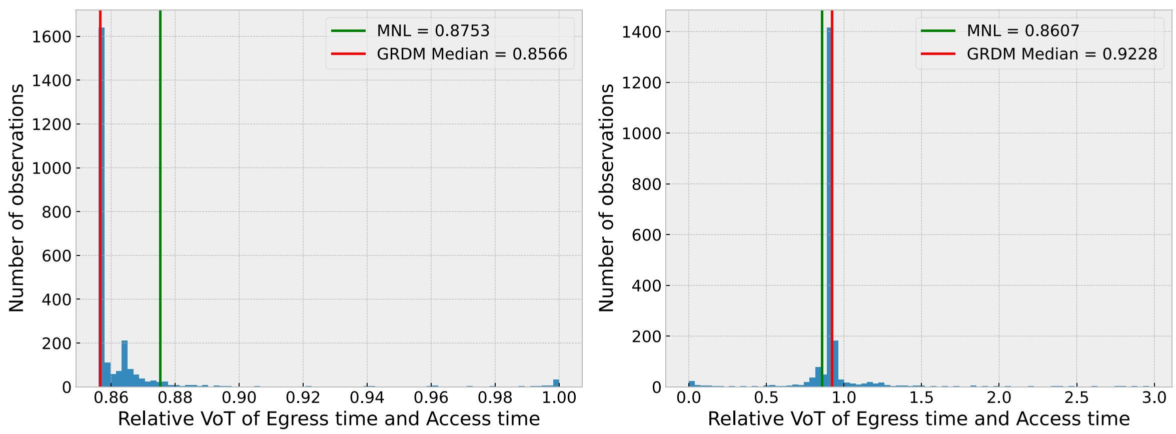

Fig. 17. Left: Distribution of the relative VoT of Egress time, work trips. Right: Distribution of the relative VoT of Egress time, leisure trips.

Table 17
Relative VoT given by each model. For the MNL + GRDM model, the median of the individual posterior relative VoT is given.

Model	MNL - Work	MNL - Lei.	MNL + GRDM - Work	MNL + GRDM - Leisure
In-vehicle time				
Bus	0.5449	0.4517	0.5232	0.4896
Local train	0.5186	0.3820	0.2908	0.1966
Metro	0.6236	0.5416	0.6286	0.5945
Reg. and intercity train	0.4923	0.4247	0.4397	0.3428
S-train	0.5799	0.4809	0.5270	0.4258
Transfer components				
Transfer penalty	5.541	5.649	4.633	5.401
Transfer walking time	0.3917	0.4989	0.3751	0.5338
Transfer waiting time	0.1028	0.0966	0.0881	0.0937
Other components				
Access time	*1*	*1*	*1*	*1*
Egress time	0.8753	0.8607	0.8566	0.9228
Half of highest headway	0.2210	0.2135	0.2642	0.2364

As Table 17 shows, the MNL and the MNL + GRDM give similar results regarding expected relative Values of Time. However, we can see that the MNL + GRDM indicates a slightly higher relative VoT for, e.g., headway, while a lower one for local train travel time.

7 Conclusion and future work

In this paper, we developed a new choice model that adopts a disjunctive decision rule, shifting from the compensatory assumption of RUMs. The new Generalised Random Disjunctive Model (GRDM) extends the existing Random Disjunctive Model (RDM) from Ehrgott et al. (2015) to allow for ranking attributes by relative importance and address issues with irrelevant attributes. We first presented the model properties and how it relates to the Universal Logit model. We then presented the formulas for the MRS, the LogSum and Elasticities. Using a LCCM, we proposed the combination of utility maximization and this disjunctive decision rule. We developed a two-class model, one using a MNL model and the other the GRDM. This combination was then tested on two large-scale case route choice studies, where we saw large improvements in fit for the combined MNL + GRDM. We gave an interpretation of the model estimates and studied the posterior MRS.

The developed GRDM offers new choice modelling insights, capturing choice behaviour that a RUM cannot explain. Its combination with a MNL in a LCCM proved effective in model fit. We tested this model and other model combinations on a bicycle route choice and a public transport route choice case study. For these cases, it outperformed by far the individual models, which supports the hypothesis that, over a series of choice situations, accounting for multiple decision rules best captures the behaviour. The MNL + GRDM LCCM outperformed the MNL + RDM LCCM, supporting the new GRDM's development. The MNL + GRDM LCCM also outperformed the MNL + MNL LCCM, which indicates that the addition of classes does not only uncover taste heterogeneity (a concern brought by Hancock and Hess (2021)) but heterogeneity in the decision rule. We also tried to disentangle further taste and decision rule heterogeneity by estimating models using a Box–Cox specification and models with many MNL latent classes. These models estimations confirmed the hypothesis that the GRDM was capturing more than taste heterogeneity or non-linear preferences.

The class allocation probabilities give interesting insights into the proportion of individuals more likely to use the disjunctive decision rule. For cyclists' route choice, it amounts to around 27.4%. For public transport leisure trips, it goes down to 11.1%, and goes further down to 6.2% for public transport work trips. These percentages can likely be attributed to the complexities of these choice tasks. Bicycle route choice is a more complicated cognitive task than public transport route choice, as the number of alternatives is much larger. Moreover, work-related trips are more likely to be repeated many times, and decision-makers are likely better to know the possible alternatives than for their leisure trips.

The GRDM also permits alternative interpretations of an individual's choice behaviour. For instance, extreme tastes for an attribute, attribute non-attendance (Swait, 2001; Hensher et al., 2005) or non-optimal choices may as well be explained by the use of a disjunctive decision rule. When studying substitution rates, individuals' non-compensatory behaviour may lead to MRS (e.g., WTD) near zero or infinity, indicating a pure non-compensatory behaviour. However, these substitution rates are choice-situation and alternative-specific and require proper posterior analysis. This model is less straightforward to interpret than traditional compensatory models, as is true for all non-compensatory models. For instance, similarly to RRM, the GRDM LogSum cannot be interpreted as a welfare metric. This model, as we saw in the cross-validation experiments, seems to have – when combined with MNL – a better forecasting ability than a model accounting for taste heterogeneity only. This model is thus particularly promising for scenario analysis and predictions. An analysis of a real-life case study confirmed the hypothesis that the GRDM is not only capturing preference heterogeneity or non-linear sensitivities but also a disjunctive behaviour that cannot be well explained by a Random Utility Model.

Future research could explore other combinations of compensatory and non-compensatory decision rules within an LCCM. The GRDM model has proved effective in capturing choice behaviour beyond utility maximization. Its performance could be compared to other non-compensatory choice models integrated within an LCCM. Moreover, a more complex RUM model could be adopted for the compensatory part of the LCCM, for example, cross nested logit (Vovsha, 1997), Mixed Logit (Revelt and Train, 1998), or a Bounded Choice Model (Watling et al., 2018). Another interesting path for future research is to investigate whether the use of decision rule is linked to any individual sociodemographic attribute.

CRediT authorship contribution statement

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

Declaration of competing interest

None.

Data availability

I have shared an example code for the new model as a GitHub link in the paper manuscript file.

Acknowledgements

We thank the Independent Research Fund Denmark (DFF) for funding this research work through the project “Next-generation route choice models for behavioural realism and large-scale applications” (Grant ID: 0136-00242B) 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 Professor Stephane Hess at the University of Leeds for giving valuable advice. Finally, we would like to thank the anonymous referees whose very relevant comments significantly improved the paper.

Appendix A. Other decision rules and examples

Conjunctive rule: For each attribute $k \in \{1, \dots, K\}$, the decision maker has a criterion χ_k that the attribute can pass or fail. Let us write $x_{ik} \sim \chi_k$ if alternative i passes χ_k. Under a conjunctive decision rule, the analyst assumes that the decision maker will select any alternative that respects the criterion for every quality.

$$y_{in} = 1 \iff \forall k \in \{1, \dots, K\},\ x_{ik} \sim \chi_k$$

Elimination-by-Aspects (EBA): Under the EBA decision rule (Tversky, 1972), The decision-maker is assumed to rank qualities by importance and to select an alternative that respects the criterion for the *main* quality. If several options remain, he/she will choose from the remaining ones according to the second quality, and so on, until one alternative remains or all qualities have been inspected. The **Lexicographic** decision rule is an example of an EBA rule, where the criterion for each quality is “being the best among the choice set”.

According to these rules, none or several alternatives may remain after applying the decision rule. If several remain, the modeller can assume several outcomes:

- The decision maker changes their criteria and applies the same decision rule to the remaining set of alternatives
- The decision maker changes their criteria and rule and applies it to the remaining set of alternatives (e.g., Utility maximization…).
- The decision maker is indifferent to all the selected alternatives and thus chooses from them with equal probabilities.

Similarly, if no alternatives remain, the decision-maker is expected to change their criteria for something less stringent.

Let us assume the route choice example from Table 1. For simplicity, we assume each attribute has the same criterion in that example.

- Under a **conjunctive** decision rule, for which the criterion is *being no more than 0.5 units worse than the best alternative*, the decision-maker chooses alternative 3.
- Under a **disjunctive** decision rule, for which the criterion is *being the best*, the decision-maker may choose alternative 1 or alternative 2.
- Under a **lexicographic** decision rule, for which the main attribute is Travel Cost, the decision-maker chooses alternative 2.
- Under an **EBA** decision rule, for which the main attributes are Travel Time, then Travel Cost, and the criterion is *being smaller than 1.5*, the decision-maker chooses alternative 3.

Appendix B. Gradient of the GRDM log-likelihood function

The gradient of the GRDM probabilities can be calculated analytically. They are useful for speeding up estimation, generating the MRS, and testing for parameter significance. Below, we describe the calculation of the hessian of the GRDM probabilities natural logarithm, as it is used in maximum likelihood estimation. For an alternative $i \in \mathscr{C}$, if we note $P_{ik} = \frac{\exp(\alpha_k x_{ik})}{\sum_{j\in\mathscr{C}} \exp(\alpha_k x_{jk})}$, we get the following derivatives, for a variance term α_l:

$$\frac{\partial \ln P_i^{GRDM}}{\partial \alpha_l} = \underbrace{\frac{\partial}{\partial \alpha_l} \ln\left(1 - \prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}\right)}_{(1)} - \underbrace{\frac{\partial}{\partial \alpha_l} \ln\left(\sum_{j\in\mathscr{C}} 1 - \prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}\right)}_{(2)}$$

The left-hand side of the equation (1) can be calculated:

$$(1) = \frac{\lambda_l \frac{\partial P_{il}}{\partial \alpha_l}(1-P_{il})^{\lambda_l - 1}\prod_{k\neq l}(1-P_{ik})^{\lambda_k}}{1 - \prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}}$$

If δ_{ij} is the Kronecker symbol, equal to 1 if $i = j$ and 0 otherwise, we have that:

$$\frac{\partial P_{il}}{\partial \alpha_l} = P_{il} \sum_{j\in\mathscr{C}} (\delta_{ij} - P_{jl}) x_{jl}$$

Thus, we have:

$$(1) = \frac{\lambda_l P_{il}\left(\sum_{p\in\mathscr{C}}(\delta_{ip} - P_{pl})x_{pl}\right)(1-P_{il})^{\lambda_l - 1}\prod_{k\neq l}(1-P_{ik})^{\lambda_k}}{1 - \prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}}$$

$$(2) = \frac{\sum_{j\in\mathscr{C}} \lambda_l P_{jl}\left(\sum_{p\in\mathscr{C}}(\delta_{jp} - P_{pl})x_{pl}\right)(1-P_{jl})^{\lambda_l - 1}\prod_{k\neq l}(1-P_{jk})^{\lambda_k}}{\sum_{j\in\mathscr{C}} 1 - \prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}}$$

We can simplify this equation multiplying the left side by $\sum_{j\in\mathscr{C}} 1 - \prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}$:

$$\frac{\partial \ln P_i^{GRDM}}{\partial \alpha_l} = \lambda_l P_i^{-1} \sum_{j\in\mathscr{C}} \left[(\delta_{ij} - P_i)\frac{P_{jl}}{1-P_{jl}}(1-P_j)\sum_{p\in\mathscr{C}}(\delta_{jp} - P_{pl})x_{pl}\right] \tag{18}$$

The differentiation of the GRDM probabilities with respect to a scale λ_l is slightly more straightforward:

$$\frac{\partial \ln P_i^{GRDM}}{\partial \lambda_l} = \underbrace{\frac{\partial}{\partial \lambda_l} \ln\left(1 - \prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}\right)} - \underbrace{\frac{\partial}{\partial \lambda_l} \ln\left(\sum_{j\in\mathscr{C}} 1 - \prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}\right)}$$

$$\frac{\partial \ln P_i^{GRDM}}{\partial \lambda_l} = \frac{\ln(1-P_{il})\prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}}{1 - \prod_{k=1}^{K}(1-P_{ik})^{\lambda_k}} - \frac{\sum_{j\in\mathscr{C}} \ln(1-P_{jl})\prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}}{\sum_{j\in\mathscr{C}} 1 - \prod_{k=1}^{K}(1-P_{jk})^{\lambda_k}}$$

This expression can be simplified similarly to Eq. (18):

$$\frac{\partial \ln P_i^{GRDM}}{\partial \lambda_l} = P_i^{-1} \sum_{j\in\mathscr{C}} (\delta_{ij} - P_i)(1-P_j)\ln(1-P_{jl}) \tag{19}$$

Appendix C. Analytical GRDM MRS and elasticities

We can analytically calculate the MRS and elasticities of the GRDM model. Keeping the same notations as Appendix B, let us remark that, for an alternative $i \in \mathscr{C}$:

$$\frac{\partial P_{il}}{\partial x_{il}} = \alpha_l P_{il}(1-P_{il})$$

and for $i \neq j \in \mathscr{C}$,

$$\frac{\partial P_{il}}{\partial x_{jl}} = -\alpha_l P_{il} P_{jl}$$

which means that, if we define δ_{ij} as the Kronecker symbol, equal to 1 if $i = j$ and 0 otherwise, we can write, for any $i, j \in \mathscr{C}$:

$$\frac{\partial P_{il}}{\partial x_{jl}} = \alpha_l P_{il}(\delta_{ij} - P_{jl})$$

We can then calculate the numerator partial derivative as follows:

$$\begin{aligned}\frac{\partial}{\partial x_{jl}}\left(1 - \prod_{k=1}^{K}\left(1-P_{ik}\right)^{\lambda_k}\right) &= \lambda_l \alpha_l P_{il}(\delta_{ij} - P_{jl})(1-P_{il})^{\lambda_l - 1}\prod_{k\neq l}\left(1-P_{ik}\right)^{\lambda_k} \\ &= \lambda_l \alpha_l \frac{P_{il}}{1-P_{il}}(\delta_{ij} - P_{jl})\prod_{k=1}^{K}\left(1-P_{ik}\right)^{\lambda_k}\end{aligned} \tag{20}$$

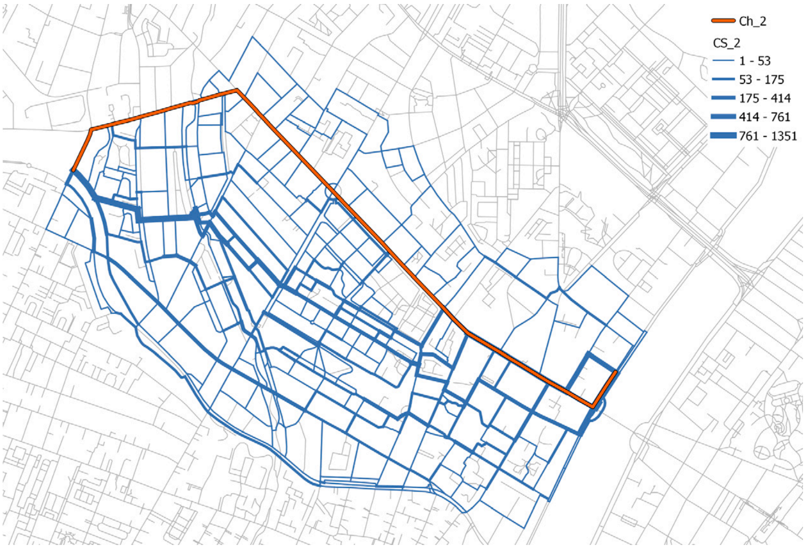

Fig. 18. Example of a choice set with the observed route. The links' width is proportional to the number of routes generated using the link (see legend in the top right corner).

It follows that:

$$\frac{\partial P_i^{GRDM}}{\partial x_{jl}} = \frac{\lambda_l \alpha_l \left[P_{il} \prod_{k=1}^{K} (1-P_{ik})^{\lambda_k} \left(\sum_{q\in\mathscr{C}} 1 - \prod_{k=1}^{K} (1-P_{qk})^{\lambda_k} \right) - \sum_{q\in\mathscr{C}} \frac{P_{ql}}{1-P_{ql}} (\delta_{iq} - P_{ql}) \prod_{k=1}^{K} \left(1-P_{qk}\right)^{\lambda_k} \left(1 - \prod_{k=1}^{K} (1-P_{ik})^{\lambda_k} \right) \right]}{\left(\sum_{q\in\mathscr{C}} 1 - \prod_{k=1}^{K} (1-P_{qk})^{\lambda_k} \right)^2}$$

$$= \lambda_l \alpha_l \sum_{q\in\mathscr{C}} (\delta_{iq} - P_i) \frac{P_{ql}}{1-P_{ql}} (1-P_q)(\delta_{qj} - P_{ql})$$

The general formula for direct and cross elasticities is given by:

$$E_{x_{jk}}^{P_i} = \frac{\partial P_i}{\partial x_{jk}} \frac{x_{jk}}{P_i} = \lambda_l \alpha_l P_i^{-1} x_{jk} \sum_{q\in\mathscr{C}} (\delta_{iq} - P_i) \frac{P_{ql}}{1-P_{ql}} (1-P_q)(\delta_{qj} - P_{ql}) \tag{21}$$

If we set $\mu_i = \ln\left(1 - \prod_{k=1}^{K} \left(1-P_{ik}\right)^{\lambda_k}\right)$, we have from Eq. (20) that ($i = j$):

$$\frac{\partial \mu_i}{\partial x_{ik}} = \lambda_k \alpha_k \frac{1-e^{\mu_i}}{e^{\mu_i}}$$

It follows that the MRS between attribute k and attribute l, for the GRDM model, are given by:

$$\mathrm{MRS}_{kl}^{i} = \frac{\frac{\partial \mu_i^{GRDM}}{\partial x_{ik}}}{\frac{\partial \mu_i^{GRDM}}{\partial x_{il}}} = \frac{\lambda_k \alpha_k P_{ik}}{\lambda_l \alpha_l P_{il}} \tag{22}$$

As the term $\frac{1-e^{\mu_i}}{e^{\mu_i}}$ simplifies. We can easily derive the MRS for the RDM as we know a GRDM is a RDM with $\lambda = (1 \ldots 1)$, so that:

$$\mathrm{MRS}_{kl}^{i} = \frac{\frac{\partial \mu_i^{RDM}}{\partial x_{ik}}}{\frac{\partial \mu_i^{RDM}}{\partial x_{il}}} = \frac{\alpha_k P_{ik}}{\alpha_l P_{il}} \tag{23}$$

These equations show that the MRS between two attributes of an alternative depend on all the other attributes (P_i is a function of all the attributes values) and all the other alternatives $j \in \mathscr{C}$.

Appendix D. Choice-set generation algorithm

The choice set generation uses a novel approach. First, it uses the stochastic approach (Nielsen, 2004; Bovy and Fiorenzo-Catalano, 2007), consisting of repeated shortest paths queries, where the link lengths are drawn randomly, following a truncated normal distribution so that negative lengths can never be drawn; $L \sim \mathscr{T}\mathscr{N}(L, \sigma L)$ with $\sigma = 0.3$. However, this method is likely to generate many highly overlapping routes because of a higher probability of a small detour becoming shorter than the actual shortest alternative than the one of a larger one. To prevent this effect, we set the number of draws to a huge number (10,000 draws for this case study). Routes were then filtered according to a local optimality criterion (Abraham et al. (2013), later used in route choice set generation by Fischer (2020)). The local optimality of a path is defined as the minimum length of a subpath of this path that is not the shortest. For instance, a route has a local optimality of 1 km if every subsection shorter than 1 km uses the shortest alternative. By setting a minimum local optimality threshold for the generated routes, we can remove routes that are not smooth (i.e., do not contain small detours) and are too similar. Indeed, two different generated routes must differ from at least the local optimality threshold. In our case, we chose a local optimality threshold of 100 m and discarded all the routes that did not respect the criterion. A minimum number of 100 routes was kept for each OD pair, even if the criterion was not respected, leading to an average choice set size of 428 routes. An example of a generated choice set is mapped in Fig. 18.

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.

Batley, R., Hess, S., 2016. Testing for regularity and stochastic transitivity using the structural parameter of nested logit. Transp. Res. B 93, 355–376. http://dx.doi.org/10.1016/j.trb.2016.07.018, Publisher: Pergamon.

Bell, D.E., 1982. Regret in decision making under uncertainty. Oper. Res. 30 (5), 961–981. http://dx.doi.org/10.1287/opre.30.5.961, Publisher: INFORMS.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, R.W. (Ed.), Handbook of Transportation Science. In: International Series in Operations Research & Management Science, Springer US, Boston, MA, pp. 5–33. http://dx.doi.org/10.1007/978-1-4615-5203-1_2.

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

Block, H.D., 1974. Random orderings and stochastic theories of responses (1960). In: Marschak, J. (Ed.), Economic Information, Decision, and Prediction: Selected Essays: Volume I Part I Economics of Decision. In: Theory and Decision Library, Springer Netherlands, Dordrecht, pp. 172–217. http://dx.doi.org/10.1007/978-94-010-9276-0_8.

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.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2), 211–252, Publisher: [Royal Statistical Society, Wiley].

Chorus, C.G., 2010. A new model of random regret minimization. Eur. J. Transp. Infrastruct. Res. 10 (2), http://dx.doi.org/10.18757/ejtir.2010.10.2.2881, Number: 2.

Chorus, C.G., 2012a. Logsums for utility-maximizers and regret-minimizers, and their relation with desirability and satisfaction. Transp. Res. Part A: Policy Pract. 46 (7), 1003–1012. http://dx.doi.org/10.1016/j.tra.2012.04.008.

Chorus, C.G., 2012b. Random regret minimization: An overview of model properties and empirical evidence. Transp. Rev. 32 (1), 75–92. http://dx.doi.org/10.1080/01441647.2011.609947.

Coombs, C.H., 1964. A Theory of Data. Wiley, Oxford, England, Pages: xviii, 585.

Dekker, T., 2014. Indifference based value of time measures for random regret minimisation models. J. Choice Model. 12, 10–20. http://dx.doi.org/10.1016/j.jocm.2014.09.001.

Dey, B.K., Anowar, S., Eluru, N., Hatzopoulou, M., 2018. Accommodating exogenous variable and decision rule heterogeneity in discrete choice models: Application to bicyclist route choice. PLoS One 13 (11), e0208309. http://dx.doi.org/10.1371/journal.pone.0208309.

Duncan, L.C., Watling, D.P., Connors, R.D., Rasmussen, T.K., Nielsen, O.A., 2020. Path size logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. Transp. Res. B 135, 1–40. http://dx.doi.org/10.1016/j.trb.2020.02.006.

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.

Ehrgott, M., Wang, J.Y.T., Watling, D.P., 2015. On multi-objective stochastic user equilibrium. Transp. Res. B 81, 704–717. http://dx.doi.org/10.1016/j.trb.2015.06.013, ISTTT 21 for the year 2015.

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

Fishburn, P.C., 1974. Lexicographic orders, utilities and decision rules: A survey. Manage. Sci. 20 (11), 1442–1471, Publisher: INFORMS.

Fosgerau, M., Łukawska, M., Paulsen, M., Rasmussen, T.K., 2023. Bikeability and the induced demand for cycling. Proc. Natl. Acad. Sci. 120 (16), e2220515120. http://dx.doi.org/10.1073/pnas.2220515120, Publisher: Proceedings of the National Academy of Sciences.

Gaundry, M.J.I., Dagenais, M.G., 1979. The dogit model. Transp. Res. B 13 (2), 105–111. http://dx.doi.org/10.1016/0191-2615(79)90028-6.

González-Valdés, F., Ortúzar, J.d.D., 2018. The stochastic satisficing model: A bounded rationality discrete choice model. J. Choice Model. 27, 74–87. http://dx.doi.org/10.1016/j.jocm.2017.11.002.

Hancock, T.O., Hess, S., 2021. What is really uncovered by mixing different model structures: Contrasts between latent class and model averaging. Eur. J. Transp. Infrastruct. Res. 21 (3), 38–63, Number: 3 Publisher: TU Delft Open.

Hancock, T.O., Hess, S., Marley, A.A., Choudhury, C.F., 2021. An accumulation of preference: Two alternative dynamic models for understanding transport choices. Transp. Res. B 149, 250–282, Publisher: Elsevier.

Hensher, D.A., Rose, J., Greene, W.H., 2005. The implications on willingness to pay of respondents ignoring specific attributes. Transportation 32 (3), 203–222. http://dx.doi.org/10.1007/s11116-004-7613-8.

Hess, S., 2012. Rethinking heterogeneity: The role of attitudes, decision rules and information processing strategies. Transp. Lett. 4 (2), 105–113. http://dx.doi.org/10.3328/TL.2012.04.02.105-113, Publisher: Taylor & Francis.

Hess, S., Daly, A., Batley, R., 2018. Revisiting consistency with random utility maximisation: Theory and implications for practical work. Theory and Decision 84 (2), 181–204. http://dx.doi.org/10.1007/s11238-017-9651-7.

Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: An approach and four case studies. Transportation 39 (3), 565–591. http://dx.doi.org/10.1007/s11116-011-9365-6.

Jedidi, K., Kohli, R., 2005. Probabilistic subset-conjunctive models for heterogeneous consumers. J. Market. Res. - J Market Res-Chicago 42, 483–494. http://dx.doi.org/10.1509/jmkr.2005.42.4.483.

Johnson, E.J., Meyer, R.J., Ghose, S., 1989. When choice models fail: Compensatory models in negatively correlated environments. J. Mar. Res. 26 (3), 255–270. http://dx.doi.org/10.2307/3172899, Publisher: American Marketing Association.

Kamakura, W.A., Russell, G.J., 1989. A probabilistic choice model for market segmentation and elasticity structure. J. Mar. Res. 26 (4), 379–390. http://dx.doi.org/10.1177/002224378902600401, Publisher: SAGE Publications Inc.

Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45 (8), 1977–1988. http://dx.doi.org/10.2307/1914121, Publisher: [Wiley, Econometric Society].

Marschak, J., 1959. Binary choice constraints on random utility indicators. Cowles Foundation Discussion Papers.

McCausland, W.J., Davis-Stober, C., Marley, A., Park, S., Brown, N., 2020. Testing the random utility hypothesis directly. Econ. J. 130 (625), 183–207. http://dx.doi.org/10.1093/ej/uez039.

McFadden, D., 1974a. Conditional logit analysis of qualitative choice behavior. Front. Econometrics Frontiers in econometrics. - New York [u.a.] : Academic Press, ISBN 0-12-776150-0. - 1974, pp. 105–142.

McFadden, D., 1974b. The measurement of urban travel demand. J. Public Econ. 3 (4), 303–328. http://dx.doi.org/10.1016/0047-2727(74)90003-6.

McFadden, D., Tye, W., Train, K., 1976. Diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. UCB-ITS-WP-7616.

Mela, C.F., Lehmann, D.R., 1995. Using fuzzy set theoretic techniques to identify preference rules from interactions in the linear model: An empirical study. Fuzzy Sets and Systems 71 (2), 165–181. http://dx.doi.org/10.1016/0165-0114(94)00266-A.

Newell, A., Simon, H.A., 1972. Human Problem Solving. h03uAAAAMAAJ, Prentice-Hall, Google-Books-ID.

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 Pract. 147, 76–92. http://dx.doi.org/10.1016/j.tra.2021.02.010.

Recker, W.W., Golob, T.F., 1979. A non-compensatory model of transportation behavior based on sequential consideration of attributes. Transp. Res. B 13 (4), 269–280. http://dx.doi.org/10.1016/0191-2615(79)90019-5.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. Rev. Econ. Stat. 80 (4), 647–657, Publisher: The MIT Press.

Sælensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecol. Econ. 59, 331–340. http://dx.doi.org/10.1016/j.ecolecon.2005.11.001.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464, Publisher: Institute of Mathematical Statistics.

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

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

Train, K.E., 2008. EM algorithms for nonparametric estimation of mixing distributions. J. Choice Model. 1 (1), 40–69. http://dx.doi.org/10.1016/S1755-5345(13)70022-8.

Tversky, A., 1969. Intransitivity of preferences. Psychol. Rev. 76 (1), 31–48. http://dx.doi.org/10.1037/h0026750, Place: US Publisher: American Psychological Association.

Tversky, A., 1972. Elimination by aspects: A theory of choice. Psychol. Rev. 79, 281–299. http://dx.doi.org/10.1037/h0032955, Place: US Publisher: American Psychological Association.

van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transp. Res. Part A: Policy Pract. 74, 91–109. http://dx.doi.org/10.1016/j.tra.2015.01.008.

Vovsha, P., 1997. Application of cross-nested logit model to mode choice in tel aviv, Israel, metropolitan area. Transp. Res. Rec. 1607 (1), 6–15. http://dx.doi.org/10.3141/1607-02, Publisher: SAGE Publications Inc.

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

Xu, Y., Zhou, J., Xu, W., 2020. Regret-based multi-objective route choice models and stochastic user equilibrium: A non-compensatory approach. Transp. A: Transp. Sci. 16 (3), 473–500. http://dx.doi.org/10.1080/23249935.2020.1719550, Publisher: Taylor & Francis.