A new mathematical representation of demand using choice-based optimization method

Meritxell Pacheco Paneque

Shadi Sharif Azadeh

Michel Bierlaire

Transport and Mobility Laboratory (EPFL) **May 2016**

STRC | **16th Swiss Transport Research Conference**

Monte Verità / Ascona, May 18 – 20, 2016

Transport and Mobility Laboratory (EPFL)

A new mathematical representation of demand using choice-based optimization method

Meritxell Pacheco Paneque, Shadi Sharif Azadeh, Michel Bierlaire
Transport and Mobility Laboratory
École Polytechnique Fédérale de Lausanne (EPFL)
Address
phone: +41-21-693-81-00
fax: +41-21-693-80-60
{meritxell.pacheco,shadi.sharifazadeh, michel.bierlaire}@epfl.ch

May 2016

Abstract

Discrete choice models are the state-of-the-art for the mathematical modeling of demand. Based on the concept of random utility, they are able to predict the choice behavior of individuals. However, these models are highly non linear and non convex in the variables of interest, and therefore difficult to be included in mixed linear optimization models. Furthermore, these models are of great importance in transportation revenue management systems. In this research, we propose a new mathematical modeling framework to include general random utility assumption inside discrete optimization framework. In order to tackle the non-linearity and non-convexity imposed by choice-models, we rely on simulation to capture the probabilistic nature of demand. Since the formulation has been designed to be linear, the price to pay is the high dimensionality of the problem. We propose an alternative formulation aiming at reducing the size of the problem. We have performed some preliminary experiments for small instances in order to compare the performances of the two models. Note that regardless from the implemented formulation, additional techniques such as decomposition methods may be required for more general instances.

Keywords
discrete choice models,combinatorial optimization, simulation

1 Introduction

During the last decade there has been an increasing trend to integrate customer behavior models in optimization. Several applications can be found in facility location problems (Haase and Müller (2014), Zhang *et al.* (2012), Benati and Hansen (2002)), and in revenue management networks in different contexts such as transportation (Haensel and Koole (2010)) and hotel management (van Ryzin and Vulcano (2014)). The main reason to combine the two is to provide a better understanding of the preferences of clients to policy makers while planning for their systems.

These preferences are formalized with predefined discrete choice models, which are the state-of-the-art for the mathematical modeling of demand. However, their complexity leads to mathematical formulations that are highly non-linear and non-convex in the variables of interest, and are therefore difficult to include in a discrete optimization model, where linearity is highly desirable, and convexity is necessary.

As a result, in the literature discrete choice models are typically assumed to be given in order to simplify the optimization model. The implicit understanding is that a complete prescription for decision problems will require fitting the right parametric choice model to data, so as to make accurate revenue predictions.

On the other hand, discrete optimization models create a platform where supply and demand closely interact, which is typically the case in transportation problems such as airline scheduling. Such models are associated with (mixed) integer optimization problems, whose discrete variables are used to design and configure the supply.

There are only few instances in the literature that have integrated discrete choice models in mixed integer linear optimization, the most typical methodology framework in operations research. Furthermore, most of them are limited to the logit model (*e.g.*, IIA assumption), where customers are assumed to be homogeneous in their observable characteristics. Many techniques have been developed in order to linearize and convexify such models. However, many of them fail to solve real cases or large instances (Azadeh *et al.* (2015)).

In this research we present a general methodology that integrates both supply and demand while keeping the discrete choice model inside the framework of a mixed linear integer problem that is scalable and solvable within reasonable time. The main objective is to incorporate state-of-the-art advanced discrete choice models within an optimization framework. With this approach, which is directly derived from the theory of utility maximization, two main issues are adressed:

- We eliminate the nonconvex representation of choice probabilitiies, which makes the optimization models computationally expensive.
- We can consider a vast class of discrete choice models, such as multivariate extreme values, latent variable and latent class models, and it is not limited just to logit models.

The paper is organized as follows. In Section 2 we introduce the framework and the demand model. A linear formulation for the maximization of revenues based on the demand is tackled in Section 3, and an alternative formulation of reduced size is described in Section 4. Finally, we use a case study from the literature to compare the performances of the two formulations in Section 5, followed by the conclusions in Section 6.

2 Demand modeling

We consider a population composed of N individuals (or groups of individuals with an homogenous behavior). The set of alternatives in the market is denoted by C, and without loss of generality it is assumed to be closed, which means that every customer chooses exactly one product. It is always possible to include an artificial "opt-out" product to capture customers leaving the market. In the considered market, each individual n has to choose one alternative within the set of her available alternatives $C_n \subseteq C$. Note that this is the first level of heterogeneity: the set of available products may vary from one customer to the next.

Discrete choice models rely on the assumption that each individual n associates a score, called *utility*, with each alternative $i \in C_n$. This utility is denoted by U_{in} and is a function of several variables describing the attributes of the alternative i and the socioeconomic characteristics of the individual n, as well as the interactions between both. The main behavioral assumption is that individual n chooses altenative i if the corresponding utility is the largest within the choice set C_n, *i.e.* if $U_{in} \geq U_{jn}\ \forall j \in C_n$. Assume there is no tie, that is for each n and $i, j \in C_n$, either $U_{in} > U_{jn}$ or $U_{jn} > U_{in}$. Then, we define the following indicator

$$w_{in} = \begin{cases} 1 & \text{if } n \text{ chooses } i, \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in C, \forall n. \tag{1}$$

Note that $w_{in} = 0$ if $i \notin C_n$.

We also define the following binary variable to account for the availability of alternatives

$$y_{in} = \begin{cases} 1 & \text{if } i \in C_n, \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in C, \forall n. \tag{2}$$

Therefore, the following statement holds for each alternative i and each individual n:

$$w_{in} = 1 \Leftrightarrow y_{in} = 1 \text{ and } U_{in} \geq U_{jn}, \forall j \in C_n. \tag{3}$$

In practice, analysts do not have access to the exact specification of the utility, and must consider it as a random variable. The most common specification is

$$U_{in} = V_{in} + \varepsilon_{in}, \tag{4}$$

where V_{in} is the deterministic part of the utility function and ε_{in} the error term, that captures everything that the analyst has not included explicitly in the model. We assume here that V_{in} is linear in the variables involved in the optimization problem, which is not necessary for the derivation of the choice model, but important in our context for its integration in the discrete optimization model. With this specification, the model becomes probabilistic and (3) is now written

$$\Pr(w_{in} = 1) = \Pr(U_{in} \geq U_{jn}, \forall j \in C_n). \tag{5}$$

It is worth nothing that the probabilistic nature of this model associates a zero probability with ties, so that they can be safely ignored, as assumed above.

By assuming a distribution for the error terms ε_{in} we can establish concrete operational models. For instance, the very well-known logit model assumes that the ε_{in} are independent (across both i and n) and identically distributed, with an Extreme Value distribution. In this case, it can be shown that (5) is characterized by

$$\Pr(w_{in} = 1) = \frac{y_{in} \exp^{V_{in}}}{\sum_{j \in C_n} y_{jn} \exp^{V_{jn}}}. \tag{6}$$

Other assumptions of the ε_{in} lead to different models, such as the *nested logit*, the *cross-nested logit* or the *logit mixtures* model, to cite just a few. Note that the formulation in (6) is non-linear as a function of the utilities and in the variables y_{in}. Some linear reformulations have been proposed in the literature (Benati and Hansen (2002), Haase and Müller (2014)).

The demand within the market for each alternative $i \in C$, understood as the number of customers choosing that alternative, can be obtained from the introduced framework. It is given by

$$D_i = \sum_{n=1}^{N} \Pr(w_{in} = 1). \tag{7}$$

3 A demand based formulation for the maximization of revenues

The demand model in (7) is in general non-linear. As stated before, various ways to linearize it have been proposed in the literature. Here we propose a different approach, which is derived directly from (4) and (5). We develop a general framework to model the demand, and we characterize its usage when the price is a decision variable of the optimization problem, the capacities for each alternative need to be taken into account and the goal is to maximize the total revenues of the operator.

3.1 A linear formulation

For each alternative i and each individual n, we rely on simulation to generate R draws $\xi_{in1},\ldots,\xi_{inR}$ from the distribution of ε_{in}. It is important to notice that this can be done for a wide variety of distributions, so that we are not restricted to logit. Each of these draws corresponds to a behavioral scenario.

Once the draws have been generated, the probabilistic nature of the model can be captured by simulation in the following way. The utility associated by individual n with alternative i, in the rth scenario is denoted by

$$U_{inr} = V_{in} + \xi_{inr} = \sum_k \beta_k x_{ink} + f(z_{in}) + \xi_{inr}. \tag{8}$$

Note that we distinguish between the part of V_{in} that is linear in the variables x_{ink}, and the part that depends on other variables z_{in}, in a possibly non linear way defined by f. The variables x_{ink} are endogeneous variables of the model, since they are those involved in the optimization problem, and z_{in} are additional exogenous variables. Then, it does not matter if f is linear or not in z_{in} because $f(z_{in})$ is a value that can be preprocessed. We also note that U_{inr} is not a random variable.

The full model, whose objective function and constraints are described below, is the following:

$$\max \sum_{i>0} R_i, \tag{9}$$

subject to

$$U_{inr} = \sum_k \beta_k x_{ink} + f(z_{in}) + \xi_{inr}, \quad \forall i, n, r, \tag{10}$$

$$y_{inr} \leq y_{in}, \quad \forall i, n, r, \tag{11}$$

$$y_{in} = 0, \quad \forall i \notin C_n, \forall n, \tag{12}$$

$$\mu_{ijnr} + \mu_{jinr} \leq 1, \quad \forall i, j, n, r, i \neq j \tag{13}$$

$$\mu_{ijnr} \leq y_{inr}, \quad \forall i, j, n, r, i \neq j \tag{14}$$

$$y_{inr} + y_{jnr} \leq 1 + \eta_{ijnr}, \quad \forall i, j, n, r, i \neq j \tag{15}$$

$$\eta_{ijnr} \leq y_{inr}, \quad \forall i, j, n, r, i \neq j \tag{16}$$

$$\eta_{ijnr} \leq y_{jnr}, \quad \forall i, j, n, r, i \neq j \tag{17}$$

$$M_{nr}\eta_{ijnr} - 2M_{nr} \leq U_{inr} - U_{jnr} - M_{nr}\mu_{ijnr}, \quad \forall i, j, n, r, \tag{18}$$

$$U_{inr} - U_{jnr} - M_{nr}\mu_{ijnr} \leq (1 - \eta_{ijnr})M_{nr}, \quad \forall i, j, n, r, \tag{19}$$

$$w_{inr} \leq \mu_{ijnr}, \quad \forall i, j, n, r, i \neq j \tag{20}$$

$$w_{inr} \leq y_{inr}, \quad \forall i, n, r, \tag{21}$$

$$\sum_{i \in C} w_{inr} = 1, \quad \forall n, r. \tag{22}$$

The revenues of alternative i are denoted by R_i (see a characterization of this quantity in the next subsection), and the objective is to maximize the sum of the revenues of all alternatives. Note that in the sum $i > 0$ is considered. This is because in addition to the I alternatives we consider $i = 0$ as the "opt-out" alternative for those customers who do not pick any alternative, either because they do not choose any alternative at all, or they choose an alternative from a competing market.

For the sake of generality, in addition to the variables y_{in} described in Section 2, we introduce the variables y_{inr} to characterize the availability of service i to individual n in scenario r. While y_{in} is a decision of the operator, that is independent of the choices of the customers, these new variables may account for the possible unavailability of an alternative due to excess of demand, as illustrated later. Constraint (11) captures the relationship between the two. We also add constraint (12) for the alternatives that are not available to individual n, as stated in (2).

Regarding the preferences of customers, we characterize the largest between the utilities of alternatives i and j. To do so, we include the indicators μ_{ijnr}, that are 1 if $U_{inr} \geq U_{jnr}$ for individual n and scenario r, and 0 otherwise. Note that it is possible that $\mu_{ijnr} = \mu_{jinr}$ if the two utilities happen to be equal, although in practice it should happen rarely. The valid inequality (13) may be included in the model to take into account the preference of only one of the two compared alternatives. We impose (14) so that alternative i cannot be preferred to j if the former is not available.

We also define the binary variable η_{ijnr} to account for the availability of alternatives i and j, since it is only necessary to compare the alternatives that are available. Then, $\eta_{ijnr} = 1$ if $y_{inr} = y_{jnr} = 1$, and 0 otherwise. In this way, if $\eta_{ijnr} = 0$, then $\mu_{ijnr}\eta_{ijnr} = 0$, so that the comparison does not make sense. This product of binary variables can be easily linearized with constraints (15), (16) and (17).

To characterize the variable μ_{ijnr} in a linear way, we define a constant M_{nr} such that

$$|U_{inr} - U_{jnr}| \leq M_{nr}, \forall i, j. \tag{23}$$

Then, it is easy to verify that constraints (18) and (19) capture the two possibilities of the definition of μ_{ijnr} by considering the following four cases:

- $\eta_{ijnr} = 1$ and $\mu_{ijnr} = 1$. Then (18) and (19) are jointly written as $0 \leq U_{inr} - U_{jnr} \leq M_{nr}, \forall i, j, n, r$. The first inequality imposes that $U_{inr} \geq U_{jnr}$, which is consistent with $\mu_{ijnr} = 1$, and the second inequality is always verified, from (23).
- $\eta_{ijnr} = 1$ and $\mu_{ijnr} = 0$. Then (18) and (19) are jointly written as $-M_{nr} \leq U_{inr} - U_{jnr} \leq 0, \forall i, j, n, r$. The first inequality is always verified, from (23), and the second imposes that $U_{inr} \leq U_{jnr}$, which is consistent with $\mu_{ijnr} = 0$.
- $\eta_{ijnr} = 0$ and $\mu_{ijnr} = 1$. Then (18) and (19) are jointly written as $-M_{nr} \leq U_{inr} - U_{jnr} \leq 2M_{nr}, \forall i, j, n, r$, and is always verified from (23).
- $\eta_{ijnr} = 0$ and $\mu_{ijnr} = 0$. Then (18) and (19) are jointly written as $-2M_{nr} \leq U_{inr} - U_{jnr} \leq M_{nr}, \forall i, j, n, r$, and is always verified from (23).

Now we need to adapt the choice variable defined in (1) to account for the choices of individuals at scenario level. We define new choice variables w_{inr}, which are 1 if individual n chooses alternative i in scenario r, and 0 otherwise. Constraint (20) states that the chosen alternative is the one with the largest utility, constraint (21) says that an available alternative is chosen, and constraint (22) imposes that exactly one choice is performed by each individual in each scenario.

With this formulation, the demand of alternative $i \in C$ within the market is given by

$$D_i = \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} w_{inr}. \tag{24}$$

The above model specification is pretty general, and is linear in the utility functions U_{inr} (and therefore in any variable appearing linearly, in particular x_{ink}), the choice variables w_{inr}, the preference variables μ_{ijnr} and the availability variables y_{in} and y_{inr}. In the next subsection we illustrate the use of this framework for instances where pricing and capacity allocation play an important role.

3.2 Dealing with prices and capacities

Consider an operator selling services to a market, where each service can be offered at a given price to a finite number of customers, called the *capacity* of the service. The demand is price elastic and heterogenous, in the sense that each group of customers may have a different behavior. Typical examples are airlines, where a service is a connection between two airports, or film distributors offering movies in various teathers.

We are aiming at finding the best strategy in terms of capacity allocation and pricing, in order to maximize the revenues of the operator. The operator is offering I services, each service i being accessible to a maximum of c_i customers, and an additional service denoted by $i = 0$ to capture the customers leaving the market. In order to consider heterogenous demand, we assume that the market is composed of N individuals, or group of individuals of homogenous behavior (in the following, we refer only to "individuals").

For the pricing strategy we consider the price as an endogenous variable in the utility function (8), that is a decision variable of the operator. The variable $p_{in} \in \mathbb{R}$ is the price that individual n must pay to access service i. Note that the index n allows the operator to propose different prices to different groups of individuals (*e.g.* students, seniors, families, etc.).

The revenues obtained by the operator from service i can be derived directly from the demand expression:

$$R_i = \frac{1}{R} \sum_{n=1}^{N} p_{in} \sum_{r=1}^{R} w_{inr}. \tag{25}$$

Since the price is an endongenous variable, (25) is non linear. A way to linearize it consists of assuming that p_{in} can only take a finite number of predetermined different values: p_{in}^1, p_{in}^2, ..., $p_{in}^{L_{in}}$, so that

$$p_{in} = \sum_{\ell=1}^{L_{in}} \lambda_{in\ell} p_{in}^{\ell}, \tag{26}$$

where $\lambda_{in\ell}$ is a binary variable that is 1 if the chosen price level is $p_{in\ell}$, and 0 otherwise. Only one price level must be selected:

$$\sum_{\ell=1}^{L_{in}} \lambda_{in\ell} = 1, \forall i, n. \tag{27}$$

The product of the binary variables $\lambda_{in\ell}$ and w_{inr} still needs to be linearized. We introduce the variables $\alpha_{inr\ell} = \lambda_{in\ell} w_{inr}$, so that

$$\lambda_{in\ell} + w_{inr} \leq 1 + \alpha_{inr\ell}, \qquad \forall i, n, r, \ell, \tag{28}$$

$$\alpha_{inr\ell} \leq \lambda_{in\ell}, \qquad \forall i, n, r, \ell, \tag{29}$$

$$\alpha_{inr\ell} \leq w_{inr}, \qquad \forall i, n, r, \ell. \tag{30}$$

Then, the expression of the revenues obtained from service i becomes

$$R_i = \frac{1}{R} \sum_{n=1}^{N} \sum_{\ell=1}^{L_{in}} \alpha_{inr\ell} p_{in}^{\ell}. \tag{31}$$

As stated before, each service i cannot accommodate more than c_i customers. If the demand for service i is larger than its capacity, a selection to decide who has access to the service and who has not must be done. Even if in a revenues maximization context the optimization algorithm will favor customers bringing the largest amount of revenues, in many situations customers arrive in a random order, and get served in a first-come-first-served basis.

Therefore, the model needs to know for each pair of individuals n and m if n has priority over m, or the other way around. A simple way to model it is to provide a priority list of individuals, where an individual is served only if all individuals before him in the list have been served. Note that the construction of this priority list can consider various aspects of the relationships between the operator and the customers, such as fidelity programs, VIP customers, etc. The priority list is supposed to be given.

In addition to constraints (27), (28), (29) and (30), the following set of constraints must be added to the model to account for capacity allocation, where $c_{min} = \min_i c_i$, $c_{max} = \max_i c_i$ and

$K_n = \max(n, c_{max})$:

$$y_{inr} \geq y_{i(n+1)r}, \qquad \forall i, n, r, \tag{32}$$

$$c_i(1 - y_{inr}) \leq \sum_{m=1}^{n-1} w_{imr} + (1 - y_{in})c_{\max}, \qquad \forall i > 0, n, r, \tag{33}$$

$$\sum_{m=1}^{n-1} w_{imr} + (1 - y_{in})c_{\max} \leq (c_i - 1)y_{inr} + K_n(1 - y_{inr}), \qquad \forall i > 0, n > c_{\min}, r, \tag{34}$$

$$\sum_{n=1}^{N} w_{inr} \leq c_i, \qquad \forall i, r. \tag{35}$$

Constraint (32) assumes that the customers are numbered according to a priority list. To verify constraint (33) we take into account the three possibilities for the availability variables (note that $y_{inr} = 1$ and $y_{in} = 0$ is not feasible due to constraint (11)):

- If $y_{inr} = 1$, then $y_{in} = 1$ (because of (11)), and this constraint becomes $0 \leq \sum_{m=1}^{n-1} w_{imr}$, that is always verified.
- If $y_{inr} = 0$ and $y_{in} = 1$, we obtain $c_i \leq \sum_{m=1}^{n-1} w_{imr}$, which means that the capacity has been reached due to the choices of individuals 1 to $n - 1$ in the priority list. It is the scenario when service i is available to n ($y_{in} = 1$), but there is not room left due to the choices of other customers ($y_{inr} = 0$).
- If $y_{inr} = 0$, then $y_{in} = 0$, and we obtain $c_i \leq \sum_{m=1}^{n-1} w_{imr} + c_{\max}$, that is always verified, as all w_{inr} are equal to 0, because service i is not available.

Finally, constraint (35) imposes that the demand cannot exceed the capacity, and to verify constraint (34) we consider the same cases for y_{in} and y_{inr}:

- If $y_{inr} = 1$, then $y_{in} = 1$ and we obtain $1 + \sum_{m=1}^{n-1} w_{imr} \leq c_i$, which imposes that the number of individuals up to and including n who have chosen i must not exceed the capacity.
- If $y_{inr} = 0$ and $y_{in} = 1$, we obtain $\sum_{m=1}^{n-1} w_{imr} \leq K_n$, that is always verified because $\sum_{m=1}^{n-1} w_{imr} \leq n \leq K_n$.
- If $y_{inr} = 0$, then $y_{in} = 0$, and we have $\sum_{m=1}^{n-1} w_{imr} + c_{max} \leq K_n$, that is always verified, as all $w_{inr} = 0$, because service i is not available. Note that no constraint is needed for individuals $n = 1, \ldots, c_{\min}$, as there is always enough capacity for these customers.

In some applications, the capacities c_i are not given and must be decided. If c_i are decision variables, the above formulation becomes non linear due to the capacity constraints (33) and (34). To linearize it we can proceed in a similar way as we have done with p_{in}. Each service is replaced by Q services, each of them with a different capacity, and such that only one will be open. We do not include here the details because we will assume for the remaining part of the paper that the capacities are given.

4 An alternative formulation

Some of the constraints in the formulation presented in Section 3 are at the order of I^2NR, when every two alternatives are compared for each individual and each scenario. This comes with a high computational price. In this section we propose a partial reformulation of the model in order to reduce its size, and potentially decrease its computational time (see Section 5 for the comparison of the performances of the two formulations).

The variables μ_{ijnr} allow us to compare the utilities pairwise. However, it is only required to identify the alternative with the highest utility among the available ones for each individual n and scenario r. In order to account for availability in this case, we define the following continuous variable:

$$\nu_{inr} = \begin{cases} U_{inr} & \text{if } y_{inr} = 1, \\ l_{inr} & \text{if } y_{inr} = 0 \end{cases} \quad \forall i \in C, \forall n, \forall r, \tag{36}$$

where $l_{inr} = \min U_{inr}$ for alternative i, individual n and scenario r.

In this way, when the alternative i is available ($y_{inr} = 1$), ν_{inr} takes the value of the utility itself (U_{inr}), but otherwise it is set to its lowest possible value, so that there might be other alternatives more "attractive" (*i.e.*with a higher utility) and "competing" to be the chosen one (*i.e.* the one with the highest utility). To characterize (36) linearly the following constraints are required:

$$l_{inr} \leq \nu_{inr}, \quad \forall i, n, r, \tag{37}$$

$$\nu_{inr} \leq l_{inr} + (m_{inr} - l_{inr})y_{inr} \quad \forall i, n, r, \tag{38}$$

$$U_{inr} + (l_{inr} - m_{inr})(1 - y_{inr}) \leq \nu_{inr}, \quad \forall i, n, r, \tag{39}$$

$$\nu_{inr} \leq U_{inr}, \quad \forall i, n, r, \tag{40}$$

where $m_{inr} = \max U_{inr}$. Constraints (37) and (38) force the variable ν_{inr} to be l_{inr} if the alternative is not available, and constraints (39) and (40) to be U_{inr} otherwise.

Note that the concept of minimum (l_{inr}) and maximum (m_{inr}) of utilities for a given alternative i, individual n and scenario r depend on the price levels (if there is more than one), according to the framework described in Section 3. In fact, all else being equal, and assuming that the β parameter associated to the price is negative (the higher the price of an alternative, the less attractive it will be for the customer), we have that the lowest price level leads to the highest utility and the other way around for the highest price level.

Once the availability at scenario level is taken into account, we define a continuous variable to capture the highest utility:

$$U_{nr} = \max_{j \in C_n} U_{jnr}. \tag{41}$$

In order to linearize the maximum of continuous variables the lower and upper bound of the variables must be provided. This is actually the case because $l_{inr} \leq U_{inr} \leq m_{inr}\ \forall i, n, r$ are known. We also need to define dummy variables to account for the alternative with the highest utility

$$\mu_{inr} = \begin{cases} 1 & \text{if } U_{nr} = U_{inr}, \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in C, \forall n, \forall r. \tag{42}$$

These binary variables are related to the availability and choice variables in the following way:

$$\mu_{inr} \leq y_{inr}, \qquad \forall i, n, r, \tag{43}$$

$$w_{inr} \leq \mu_{inr} \qquad \forall i, n, r, \tag{44}$$

where constraint (43) states that an unavailable alternative cannot be the one with the highest utility, and constraint (44) says that an alternative that does not have the highest utility cannot be chosen.

Finally, the constraints for the linearization of the U_{nr} are

$$\nu_{inr} \leq U_{nr}, \qquad \forall i, n, r, \tag{45}$$

$$U_{nr} \leq \nu_{inr} + M_{inr}(1 - \mu_{inr}) \qquad \forall i, n, r, \tag{46}$$

$$\sum_{i \in C} \mu_{inr} = 1, \qquad \forall n, r, \tag{47}$$

where $M_{inr} = \max_{j \in C_n} m_{jnr} - l_{inr}$. Note that we consider ν_{inr} instead of U_{inr} itself to account for the availability of alternative i, as described above. We consider two cases:

- $y_{inr} = 1$. Then $\nu_{inr} = U_{inr}$ and constraint (45) is written as $U_{inr} \leq U_{nr}$, which is always satisfied. Constraint (46) has two possibilities according to the value of μ_{inr}:
  - $\mu_{inr} = 0$ means that there is an alternative with a higher utility, so this constraint is written as $U_{nr} \leq U_{inr} + M_{inr}$ and is always verified.
  - $\mu_{inr} = 1$ means that alternative i has the highest utility, that is $U_{nr} = U_{inr}$. This equality is actually obtained from constraints (46), which is written as $U_{nr} \leq U_{inr}$, and constraint (45), that contains the other sense of the inequality.

- $y_{inr} = 0$. Then $v_{inr} = l_{inr}$ and constraint (45) is written as $l_{inr} \leq U_{nr}$, that is always satisfied. According to constraint (43), μ_{inr} can only be 0. Then, constraint (46) is written as $U_{nr} \leq l_{inr} + M_{inr} \leq \cancel{l_{inr}} + \max_{j \in C_n} m_{jnr} - \cancel{l_{inr}}$, which also holds since $U_{inr} \leq m_{inr}$ $\forall i \in C_n, n, r$.

The rest of the framework described in Section 3 remains the same. With this new formulation there are no variables at order I^2NR anymore, being the highest order INR. In the following section a real case study from the literature is considered in order to compare the computational times and the objective values of the two models.

5 Preliminary experimental results

For the comparison of performances we consider the case study developed in Ibeas *et al.* (2014), whose data was kindly provided by the authors. Their goal is to assess the local authorities of a small coastal town of Spain for the building of an underground car park due to the lack of available parking. They use a mixed logit model (*i.e.* accounting for random taste parameters) to characterize the behaviour of potential car park users when choosing a parking place. We have chosen this work because it allows us to illustrate our formulation for models different than logit.

The data was collected from a stated preferences survey, with a final sample size of 197 respondents. The choice set was composed by three alternatives: the two existing ones, free on-street parking (FSP) and paid on-street parking (PSP); and the one to be tested, paid parking in an underground car park (PUP). The survey was composed of eight choice scenarios, each of them with different values for the variables considered in the experimental design: access time to the parking, access time to the destination and parking fee.

Our fomulation is not designed to take into account stated preferences data, so we need to adapt the dataset from the case study to our particular case. To do so we consider only one of the scenarios (scenario 1) to characterize the values of access time to the parking and access time to the destination for the three alternatives. Note that parking fee is not an input value. In fact, it is the only decision variable that we consider in the optimization problem. Its levels are those defined in the different scenarios. In particular, 0.6 and 0.8 for FSP, and 0.8 and 1.5 for PUP. Note that FSP has no price levels because it is free, so it is always 0. Finally, since the capacities for the alternatives are not specified, we can set them to any value to generate different instances.

The utility functions for each alternative are described by the author. They include as attributes

Table 1: Performance of the two models for a subset of 25 individuals

		Capacities			Initial model		Alternative model		
N	R	c_{FSP}	c_{PSP}	c_{PUP}	Time(s)	Obj	Time(s)	Obj	Gap (%)
25	1	10	10	10	0.37	18.30	0.20	18.30	45.95
25	5	10	10	10	6.48	18.58	3.20	18.58	50.62
25	10	10	10	10	32.65	18.86	8.49	18.86	74.00
25	50	10	10	10	353.47	18.89	74.21	18.89	79.01
25	100	10	10	10	1060.32	18.92	431.46	18.92	59.31

Source: own

Table 2: Performance of the two models for a subset of 50 individuals

		Capacities			Initial model		Alternative model		
N	R	c_{FSP}	c_{PSP}	c_{PUP}	Time(s)	Obj	Time(s)	Obj	Gap (%)
50	1	20	20	20	0.58	33.10	0.43	33.10	25.86
50	5	20	20	20	55.95	32.26	11.58	32.26	79.30
50	10	20	20	20	307.95	31.56	97.12	31.56	68.46
50	25	20	20	20	1616.57	32.23	763.37	32.23	52.78

Source: own

of the alternatives the variables of the experimental design (for the case study the parking fee was given by the configuration of the scenario), being the access time to the parking and the parking fee the ones with random parameters (different among individuals). Regarding the socioeconomic characteristics of the individuals, the origin of the journey (if it is internal or external to the twon, only for FSP) and the age of the vehicle (only for PUP) are taken into account. They also consider the interactions of parking fee with socioeconomic variables, in particular low income level and the fact of being resident in the study area. The utility of the "opt-out" alternative is defined in a way that it is the less atractive for all individuals in all scenarios.

The optimization problem has been implemented in a C++ environment, and has been solved with CPLEX 12.6.2. For the sake of illustration, and given the large dimension of the model, it has been solved for a reduced subset of individuals and a reduced number of draws. Tables 1 and 2 show the values that have been set for the capacities (that have been set in such a way that they are not too tight but also not too loose), the computational times and the objective values for the two formulations, and the gap between the times calculated as

$$\frac{t_{alternative} - t_{initial}}{t_{initial}} \cdot 100. \tag{48}$$

In both cases, the computational time is improved substantially, being in some cases decreased by more than 70%. The objective values are the same in both formulations. We note that the values that are supposed to be randomly generated in the model are the same for the two instances, that is they have been generated beforehand. Regarding the number of draws, the idea is to consider a number of draws such that the objective value stabilizes. In this case, it keeps growing for 25 indivdiuals and oscillating for 50 individuals, which shows that more draws may be required.

6 Conclusions

In this paper, we presented a new mathematical model that integrates choice modeling with optimization in a linear way. In fact, with the help of utility maximization theory and simulation, we suceed to overcome the non-linearity and non-convexity caused by the choice probabilities. Furthermore, we also introduced an alternative equivalent formulation of reduced size that speeds up the computational time, as shown by the results.

Despite the improvement, for a population of 50 individuals, a run with only 25 draws already takes almost 30 minutes, and we are interested in a higher number of draws, since the more draws we consider, the better the estimation of the objective value. Then, when the number of alternatives, simulation draws and individuals grow, the problem might take really long to be solved. However, since the individuals and the scenarios are independent from one another, decomposition techniques may be considered to solve large instances.

Acknowledgement

We are grateful to the authors in Ibeas *et al.* (2014) for providing us the data that allowed us to test the two formulations.

7 References

Azadeh, S. S., P. Marcotte and G. Savard (2015) A non-parametric approach to demand forecasting in revenue management, *Computers & Operations Research*, **63**, 23–31.

Benati, S. and P. Hansen (2002) The maximum capture problem with random utilities: Problem formulation and algorithms, *European Journal of Operational Research*, **143**, 518–530.

Haase, K. and S. Müller (2014) A comparison of linear reformulations for multinomial logit choice probabilities in facility location models, *European Journal of Operational Research*, **232** (3) 689–691.

Haase, K. and S. Müller (2014) A comparison of linear reformulations for multinomial logit choice probabilities in facility location models, *European Journal of Operational Research*, **232**, 689–691.

Haensel, A. and G. Koole (2010) Estimating unconstrained demand rate functions using customer choice sets, *Journal of Revenue & Pricing Management*, **10** (5) 438–454.

Ibeas, A., L. dell'Olio, M. Bordagaray and J. de D. Ortúzar (2014) Modelling parking choices considering user heterogeneity, *Transportation Research Part A: Policy and Practice*, **70**, 41 – 49, ISSN 0965-8564.

van Ryzin, G. and G. Vulcano (2014) A market discovery algorithm to estimate a general class of nonparametric choice models, *Management Science*, **61** (2) 281–300.

Zhang, Y., O. Berman and V. Verter (2012) The impact of client choice on preventive healthcare facility network design., *OR Spectrum*, **34**, 349–370.