Integrating advanced discrete choice models in mixed integer linear optimization

Meritxell Pacheco Paneque [*] Shadi Sharif Azadeh [†] Michel Bierlaire [*]
Bernard Gendron [‡]

July 14, 2017

[*]Transport and Mobility Laboratory (TRANSP-OR), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, {meritxell.pacheco, michel.bierlaire}@epfl.ch

[†]Operations Research & Logistics group, Econometric department, School of Economics Erasmus University, Rotterdam, The Netherlands, sharifazadeh@ese.eur.nl

[‡]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada & CIRRELT, bernard.gendron@cirrelt.ca

Abstract

The integration of customer behavioral models in operations research (OR) is appealing to operators and policy makers (the supply) because it provides a better understanding of the preferences of customers (the demand) while planning for their systems. These preferences are formalized with discrete choice models, which are the state-of-the-art for the mathematical modeling of demand, whereas mixed integer linear programming (MILP) models are considered to design and configure the systems. Notwithstanding the clear advantages of this integration, the complexity of discrete choice models leads to mathematical formulations that are highly nonlinear and nonconvex in the variables of interest, and therefore difficult to be included in MILP. In this paper, we present a general framework that overcomes these limitations by integrating in MILP advanced discrete choice models. A concrete application on benefit maximization from an operator selling services to a market is used to illustrate the employment of the framework. A case study from the recent literature is considered to perform various experiments, such as price differentiation by population segmentation. The results show that this approach is a powerful tool to configure systems based on the heterogeneous behavior of customers, and it allows to investigate advanced marketing strategies and business models.

Keywords: mixed integer linear programming; discrete choice models; customer behavior; simulation; combinatorial optimization; supply-demand interaction

1 Introduction

The success stories of operations research (OR) are countless. The tremendous development of advanced mathematical formulations, and of efficient algorithms, able to solve problems of gigantic sizes, allow to deal with an impressive spectrum of applications in transportation, finance, business, health, manufacturing, and supply chain, to name just a few.

Interestingly, the vast majority of OR models found in the literature focus almost exclusively on the "supply" side of the problem. The "demand" side is often neglected, assumed given, or modeled in a simple way (see the discussion in Section 2).

Since the Nobel Prize in Economics was awarded to Daniel McFadden in 2000, the research on discrete choice models has also experienced a tremendous development. Advanced disaggregate behavioral models, predicting the choice of individuals or groups of individuals, accounting for subjective aspects such as attitudes and perceptions, have successfully been specified, calibrated from data, validated and applied to predict the demand in a great deal of contexts.

Although the "supply" and the "demand" closely interact in the real world, the two fields of research have developed independently, with little interaction between the two. The main reason of this lack of interaction is certainly the completely different types of focus of each field. In OR, the tractability of mathematical formulations requires the linearity or the convexity of the involved functions. Mixed integer linear programming (MILP) formulations represent a significant share of the models reported in the literature. In discrete choice, the focus is on behavioral realism, building on the micro-economic theory of utility. The methodological developments in this field are mainly about the relaxation of unrealistic assumptions in order to better reproduce the actual behavior. As a consequence, the mathematical formulations are complex, and certainly not linear or even convex. They are therefore not appealing for the OR modelers.

The objective of this paper is to bridge this gap. We propose a general framework that allows to integrate in a MILP formulation a choice model based on the random utility principle. The only condition that is required to obtain a MILP model is that the variables appearing both in the demand and the supply side of the problem appear linearly in the utility function.

The key idea is to rely on simulation to linearize the choice model. The main consequence is the potentially large size of the resulting formulations. However, the advantage of simulation is that the trade-off between model accuracy and tractability can be explicitly controlled by the modeler. Moreover, recent advances in decomposition methods, particularly well suited for this framework, allow to handle MILP models of very large sizes.

The remainder of the paper is organized as follows. Section 2 includes several references from the literature where demand and supply closely interact. Section 3 describes the general framework and characterizes the mathematical model. Section 4 contains a concrete application to illustrate the use of the formulation, and the case study used for the proof of concept is detailed in Section 5. Finally, the conclusions and future work are discussed in Section 6.

2 Literature review

For many problems addressed within the scope of OR, the demand-supply tradeoff plays an important role. Facility location, revenue management, transportation, supply chain management and logistics are some general topics in which the demand assumptions are a key element when it comes to the complexity and reliability of the associated mathematical models.

In the context of facility location, the difficulties associated with the decisions on spatial resource allocation lead to complicated formulations (Laporte et al.; 2015). This has historically limited the research to deterministic problems, where all input parameters (including the demand) are considered as known quantities. Since in reality these parameters may broadly fluctuate, researchers have focused on the development of formulations capturing this uncertainty. In these models, demand is generally assumed to follow a probability distribution or to change its patterns under different hypothetical scenarios. In addition to these modeling strategies, and even if conveying to inaccurate demand estimates, simplifications such as aggregating demand points are commonly employed (Snyder; 2006, Francis and Lowe; 2015).

Demand forecasting is also a critical aspect in revenue management (RM) systems because of its direct influence on the booking limits, which determine the profits. Price is an essential factor in operational planning, as it is one of the most effective variables that managers can manipulate to encourage or discourage demand in the short run (Bitran and Caldentey; 2003, Sharif Azadeh et al.; 2014). Despite its relevance, in the immense majority of RM systems (such as airlines or retail), demand is assumed to be independent, i.e., isolated from its market environment (van Ryzin; 2005). Furthermore, the lack of information about customers' preferences, as well as the complexity of the resulting mathematical formulations, make disaggregate forecasting extremely difficult and infrequently used in practice (Talluri and Van Ryzin; 2006).

Understanding the underlying behavior of individuals and incorporating it into the modeling framework are equally crucial in other application areas. For instance, in transportation, some travelers may modify their travel arrangements (e.g., departure time, route) depending on the level of service of the network. In supply chain management, appropriate supply responses based on customers' reactions need to be provided in case of stockouts. The discussed application areas represent a small collection of examples illustrating the necessity of a better representation of the demand that accounts for the heterogenous behavior of users by means of modeling the demand (either deterministic or stochastic) in a disaggregate way, so that the individuals constitute the fundamental unit of demand. To address this issue, behavioral models capturing the heterogeneity of the demand are required.

The most advanced and operational behavioral models are choice models. First introduced by McFadden (1974), they are able to predict the choice behavior of individuals in detail, taking into account not only attributes of the goods, such as price or quality, but also socio-economic characteristics of the individuals, such as age or income. Consequently, they allow to capture the heterogeneity of the behavioral patterns in the population, which in turn generate the demand.

The integration of choice models in optimization problems is an increasing trend. Besides accounting for the demand heterogeneity, these models enable to include other features of the demand, such as complex substitution patterns, and to investigate other phenomena, such as demand elasticity. Recent literature can be found in the application areas mentioned above.

In facility location, Benati (1999) and Benati and Hansen (2002) address the problem of a firm that wants to enter a competitive market by locating p new facilities in order to maximize its market share. Since the firm cannot predict customers' behavior deterministically, a logit model is employed. In the former paper, the resulting optimization model is reformulated as a p-median problem and it is solved by Lagrangian relaxation and branch and bound. In the latter, the model is a hyperbolic sum integer optimization problem that proves to be NP-hard. The authors develop three solution methods (one exploiting the concavity of the objective function and the other two reformulating the problem as a MILP model), and they show that only moderate size problems can be solved up to optimality.

A similar problem is considered in Marianov et al. (2008). They assume that customers decide what facility to patronize based on the travel and waiting times by means of a logit model, which results in a nonlinear integer problem. Due to its complexity, a heuristic method is proposed. In Haase (2009), two new models for discrete location planning under static competition are defined. The first one considers the IIA property of the basic logit model (i.e., constant substitution patterns among alternatives), whereas the second one relaxes this property (i.e., it allows for flexible substitution patterns). An analogous model is presented in Aros-Vera et al. (2013). In both cases, the probability equation given by the logit model is replaced by an equivalent set of constraints that transforms the formulation into a MILP model. Again, a heuristic is proposed to solve large instances.

In the context of school location with free school choice, Müller et al. (2009) define a two-step approach in order to minimize the location and transportation costs over a given time horizon with respect to students choosing the school with the highest utility. As in a spatial context, the IIA property of the logit is unlikely to hold. A mixture of logit models, which allows for flexible substitution patterns and random taste variation among individuals, is applied. The first step allocates students for each *scenario* (combination of open and closed schools) according to capacity and utility, and in the second step, a scenario is selected for each period, while minimizing the total costs. Haase and Müller (2013) propose a MILP model where the objective is to maximize the standardized expected utility of all students, whose values are simulated with a mixture of logit models. They show that real problems can be solved optimally (or closed to optimality) within a few minutes with state-of-the-art solvers.

In Zhang et al. (2012), two alternative models for designing a service facility network are introduced: a "probabilistic-choice model," in which the logit probabilities to patronize each facility (including the option of not patronizing any facility) are characterized, and an "optimal-choice model," which specifies that each customer will go to the most attractive facility. Both problems are formulated as MILP models, and are solved to optimality with standard MILP solvers for small- and medium-size instances, and with heuristics for large instances. This linear reformulation for the logit probabilities is compared in Haase and Müller (2014) with the linearizations defined in Benati and Hansen (2002) and Haase (2009), concluding that the approach in Haase (2009) seems to be promising for large problems.

Customer-behavior-oriented models of demand represent a promising approach for RM, and research on choice models applied to RM problems is advancing rapidly (van Ryzin; 2005, Shen and Su; 2007). These models were first introduced in Andersson (1998), where a logit model is assumed to compute the probability of a passenger that was rejected at one flight-class combination to request a seat at another flight-class, called the recapture rate (or buy-up rate). An example is considered to show that an increase in the revenue is experienced when implementing recapture and buy-up.

In Talluri and Van Ryzin (2004), the authors provide a general analysis of the impact of choice behavior in RM. In particular, they explicitly model consumer choice behavior using a general choice model where the probabilities of purchasing each fare product depend on the set of available fare products. They test their method against the traditional ones, and show that significant improvements in the revenue can be achieved with choice-based RM models. Similar results are obtained in Vulcano et al. (2010), where the authors rely on a multisegment (i.e., customers belonging to discrete segments) logit model. As discussed by the authors, other approaches such as nested logit models might be more appropriate, but at the same time more difficult to tackle computationally. In Sharif Azadeh et al. (2015), a non-parametric approach for demand forecasting is defined. The authors propose a mixed integer nonlinear programming (MINLP) model to estimate product utilities, as well as to capture seasonal effects. The MINLP model is linearized with local convexification and relaxation techniques, and solved using global optimization by introducing a tailored branch-and-bound algorithm.

In Liu and Van Ryzin (2008), the analysis performed in Talluri and Van Ryzin (2004) is extended to the network setting. The authors formulate a general model of RM under customer choice behavior, for which they characterize the corresponding choice-based deterministic linear programming model (CDLP), as proposed in Gallego et al. (2004) (which can be considered as a deterministic approximation of the original stochastic problem). The authors show that the performance of the CDLP is asymptotically optimal as capacity and demand are scaled up. Furthermore, they develop a decomposition heuristic to convert the static CDLP solution into a dynamic control policy. Even if the CDLP and the decomposition heuristic are computationally complex, they show that under logit with disjoint segments model (i.e., customers divided into segments, each of them having a disjoint choice set) both can be solved efficiently.

The problem of setting profit maximizing tolls over a subset of arcs of a transportation network where the users minimize a disutility function (comprising the fixed costs and tolls from their origin to their destination) is addressed in Gilbert et al. (2014a) and Gilbert et al. (2014b). In the former, a logit route choice model is assumed to account for users' awareness of the network conditions. The resulting optimization problem is nonlinear and nonconvex, and may have several local optima. In the latter, a mixture of logit models (i.e., price sensitivity distributed across users) is assumed. In contrast with simpler random utility models, no closed form solution is available for the assignment of users to paths of the transportation network, which makes this framework numerically challenging.

In the discussed examples, the probabilistic representation of the choice is either included in a deterministic way, i.e., the utility is considered as exogenous to the optimization model, or the decision variables of the optimization problem appear in the utility function, i.e., the utility is endogenous to the optimization model. The second approach is obviously more challenging, since it leads to nonlinear and nonconvex formulations, but it allows for capturing the interaction between demand and supply. Furthermore, in order to come up with tractable and more efficient solutions, various authors have placed simplistic assumptions on the choice model, which might be inappropriate in reality (Vulcano et al.; 2010, Liu and Van Ryzin; 2008). However, more advanced demand models, based on mixtures of logit (such as hybrid choice models), even if they have shown to better forecast the behavior of individuals, have no closed-form, and are therefore difficult to integrate into optimization models.

The review of the literature illustrates that the momentum for using choice models into optimization models is building. Nevertheless, the capability of choice models to predict heterogeneous and disaggregate demand is by far not enough exploited in the OR literature. In this paper, we propose a mathematical formulation that is designed to integrate discrete choice models inside MILP. The formulation is linear, in order to ensure the tractability of the optimization model, but remains fairly general, in the sense that the framework can be employed with any choice model.

3 General framework

The framework aims at providing a mathematical representation of a behavioral model of the demand, which can be integrated into an optimization model characterizing the supply.

The *demand* is the result of the choice of many individuals, who we call "customers": the choice of buying a given product, the choice of traveling with a given mode of transportation, the choice of going to a specific movie theater, etc. Such a representation of demand is referred to as "disaggregate." It allows to account for the heterogeneity of the population of interest, where customers have different tastes and preferences. The choice itself is captured by a discrete choice model that predicts the choice of each customer from a finite set of discrete alternatives.

The *supply* is characterized by the decisions of the entity in charge of the configuration of the system, who we call the "operator": the price of a product, the schedule of a public transportation service, the type of movie that is displayed in a theater, etc. The typical focus of OR is to optimize the supply.

In the modeling framework that we are describing, the demand and the supply are characterized by variables of three types:

1. the exogenous demand variables $x^d \in \mathbb{R}^D$,
2. the exogenous supply variables $x^s \in \mathbb{R}^S$, and
3. the endogenous variables $x^e \in \mathbb{R}^E$, involved both in the demand and the supply representations.

As discussed below, these variables can be restricted to take integer or binary values, depending on the specific model. The exogenous variables appear either in the demand model or in the supply model, but not in both. The endogenous variables are present in both, and characterize the interactions between demand and supply.

A typical example of an endogenous variable is the price of a service. The operator decides on a price for a service in order to maximize its revenue, and the customer reacts to the price in order to decide if she buys the service or not. If the operator sets a price that is too high, few customers will access the service, and a low revenue will be generated. If the price is too low, many customers will use the service, but the generated revenue will also be low. This example is treated extensively in Section 5. Other examples of endogenous variables are the schedule of an event (e.g., opening hours of a shop, departure of a train, show time of a performance) and the capacity of a facility.

3.1 The supply model

We assume that the supply model is composed of:

- an objective function $g^s : \mathbb{R}^{S+E} \to \mathbb{R}$ that relates the decisions at the supply level to an aggregate performance of the system:
$$g^s(x^s, x^e) \tag{1}$$
- a set of constraints that identifies the feasible configurations of the variables:
$$h^s(x^s, x^e) \leq 0, \tag{2}$$
$$l^e \leq x^e \leq m^e, \tag{3}$$
where $h^s : \mathbb{R}^{S+E} \to \mathbb{R}^I$, $I \geq 0$ is the number of constraints, $\ell^e \in \mathbb{R}^E$ is the vector of lower bounds on x^e and $m^e \in \mathbb{R}^E$ is the vector of upper bounds.

Any typical OR model fits this general representation. In the following, it is assumed that the functions g^s and h^s are linear in the variables of interest in order to derive a MILP formulation.

To illustrate the idea, consider the case of a company running movie theaters that needs to decide what movies to schedule at what time, in what theater, and at what price. The show time, the choice of the movie and the price are endogenous variables. Decisions about staff and equipments are exogenous variables. The objective function g^s computes the total benefit, calculated as the difference between the total revenue generated by the ticket sales and the operational costs. The constraints h^s may include the fact that two movies cannot be displayed in the same theater at the same time, that a given theater cannot fit more spectators than the number of seats, etc.

3.2 The demand model

Regarding the demand model, the set of all potential alternatives is called the *choice set* and is denoted by $\mathcal{C}$. The alternatives in $\mathcal{C}$ are indexed by i. We consider a population of N customers, indexed by $n \geq 1$. Generally, it is impossible to have access to the full population, and a sample must be used. The following description, based on the full population, can be easily adapted to a representative sample.

Note that the choice set of two different customers may not be the same. The choice set of customer n is denoted by $\mathcal{C}_n \subseteq \mathcal{C}$. For instance, some people just do not like action movies, and are therefore not even considering them. These decisions are modeled with the following binary variables:

$$y^d_{in} = \begin{cases} 1 & \text{if } i \in \mathcal{C}_n, \\ 0 & \text{otherwise,} \end{cases} \quad \forall i, n. \tag{4}$$

These variables are the result of exogenous demand decisions that are known in advance.

Also, some alternatives may not be offered by the operator for certain reasons. For example, from a benefit maximization point of view, a movie that is not profitable will not be proposed (see Section 4). These decisions are endogenous and are modeled with the binary variables y_i, which are 1 if alternative i is offered to the customers and 0 otherwise. The endogenous variables y_i belong to the vector x^e. We define the variables y_{in} as the product of both decisions, i.e.,

$$y_{in} = y^d_{in} y_i. \tag{5}$$

These variables are equal to 1 when customer n considers alternative i and this is proposed by the operator, and 0 otherwise.

For each alternative i, we denote by c_i its capacity, that is, the maximum number of customers who can choose it. In the case of the movie theater, this is typically the number of seats available in a given theater for a given show time.

The choice of customers is modeled using a *discrete choice model*. These models rely on the assumption that each customer n associates a score, called *utility*, with each alternative $i \in \mathcal{C}_n$. This utility is defined as

$$U_{in}(x^d, x^e), \tag{6}$$

where $U_{in} : \mathbb{R}^{D+E} \to \mathbb{R}$ is a function of the demand variables x^d and x^e. The behavioral assumption is that customer n chooses alternative i if the associated utility is the largest within the choice set $\mathcal{C}_n$, i.e., if

$$U_{in} \geq U_{jn}, \forall j \in \mathcal{C}_n. \tag{7}$$

It is also assumed that each customer chooses one and only one alternative.

In practice, the analyst does not have access to the complete specification of the utility function U_{in}. Thus, utility is modeled as a random variable $U_{in}(x^d, x^e; \varepsilon_{in})$, where ε_{in} is a random variable. The most common specification involves an additive error term:

$$U_{in}(x^d, x^e; \varepsilon_{in}) = V_{in}(x^d, x^e) + \varepsilon_{in}, \tag{8}$$

where $V_{in} : \mathbb{R}^{D+E} \to \mathbb{R}$ is the deterministic part of the utility function, that includes everything that can be modeled by the analyst, and ε_{in} is the error term, that captures everything that has not been included explicitly in the model and is independent of the variables x^d and x^e. Operational choice models are obtained by assuming a distribution for ε_{in}. For example, the *logit model* is obtained by assuming that ε_{in} are independent and identically distributed (across both i and n), with an extreme value distribution. Other assumptions lead to different models, such as the nested logit, the cross nested logit or the mixtures of logit models, to cite a few.

In the case study described in Section 5, a mixture of logit models is considered. In this case, the deterministic part of the utility specification of the standard logit model (generally $V_{in} = \beta x_{in}$, where β is the vector of coefficients and x_{in} the demand variables for alternative i and customer n) is generalized by allowing the coefficients to be randomly distributed among the customers, i.e., by assuming β_n instead of β, where $\beta_n \sim f(\beta|\theta)$, θ being the parameters of the distribution of β_n, such as their mean and variance.

The probability that customer n chooses alternative i within the choice set $\mathcal{C}_n$ is

$$P_n(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}, \forall j \in \mathcal{C}_n). \tag{9}$$

In the following, it is assumed that U_{in} is linear in the endogenous variables. This is not required as such for the derivation of the choice model, but important in our context for its integration in a MILP formulation. For instance, the deterministic term in (8) can be written as

$$V_{in}(x^d, x^e) = \sum_k \beta_k x^e_{ink} + q^d(x^d), \tag{10}$$

where x^e_{ink} is the kth endogenous variable associated by customer n with alternative i and x^d are the exogenous demand variables. The function $q^d : \mathbb{R}^D \to \mathbb{R}$ does not need to be linear since these variables are not involved in the supply side, and can be calculated separately.

In the case of the logit model, it can be shown that (9) is written as

$$P_n(i|x^d, x^e, y_i) = \frac{y_{in} e^{V_{in}(x^d, x^e)}}{\sum_{j \in \mathcal{C}} y_{jn} e^{V_{jn}(x^d, x^e)}}. \tag{11}$$

Note that the formulation (11) is nonlinear as a function of the endogenous variables x^e and y_i.

In the case of the mixture of logit models, the probability that customer n chooses service i is given by the standard logit formula conditional on β_n:

$$\Gamma(i|x^d, x^e, y_i; \beta_n) = \frac{y_{in} e^{V_{in}(x^d, x^e; \beta_n)}}{\sum_{j \in \mathcal{C}} y_{jn} e^{V_{jn}(x^d, x^e; \beta_n)}}. \tag{12}$$

As β_n is random and unknown, the (unconditional) choice probability (9) is the integral of the logit formula over the density of β_n:

$$P_n(i|x^d, x^e, y_i) = \int \Gamma(i|x^d, x^e; \beta) f(\beta|\theta) d\beta. \tag{13}$$

The expected demand for each alternative $i \in \mathcal{C}$ is then given by

$$D_i = \sum_{n=1}^{N} P_n(i|x^d, x^e, y_i). \tag{14}$$

In summary, we assume that the following is provided as an input for the supply model:

- the values of the exogenous supply variables x^s,
- the functions g^s and h^s, characterizing the optimization problem,

and the following for the demand model:

- the values of the exogenous demand variables x^d,
- the universal choice set $\mathcal{C}$,
- the size of the population N,
- the values of the availability variables y^d,
- the capacities c_i, and
- the function q^d.

3.3 Linearization through simulation

In this section, we derive a formulation of the demand model that is linear in its endogenous variables, so that it can be integrated in a MILP model. The nonlinearity of the demand model (14) is due to the random nature of the utility function. In order to circumvent it, we rely on simulation. For each ε_{in} in (8), we generate R draws $\xi_{in1},\ldots,\xi_{inR}$ based on the distributional assumption. For example, if the model is logit, we generate draws from an extreme value distribution.

Once the draws have been generated, for each scenario r we obtain the utility associated with alternative i by customer n. For the specification (10), we have

$$U_{inr} = V_{in} + \xi_{inr} = \sum_{k} \beta_k x^e_{ink} + q^d(x^d) + \xi_{inr}. \tag{15}$$

As the variables x^e_{ink} are bounded (see (3)), and the variables x^d are given, lower and upper bounds on U_{inr}, denoted by ℓ_{inr} and m_{inr}, can be derived:

$$\ell_{inr} \leq U_{inr} \leq m_{inr}. \tag{16}$$

Availability of alternatives An alternative may be unavailable for three reasons. First, an alternative might not be considered by the customer ($y^d_{in} = 0$). Second, the operator decides that the alternative is not made available ($y_i = 0$). Third, the alternative may be unavailable because its capacity has already been reached. In the case of the movie theater, the movie that a certain customer would like to watch is proposed by the theater, but is fully booked. This type of unavailability is more complex to model, as it is not a direct decision as such, but the result of the decisions of other customers. Note that, in our framework, this can vary from one scenario to the next. Indeed, an alternative might be attractive in one scenario, generating more demand than its capacity, and less attractive in another.

We model the availability of alternative i to customer n in scenario r using the binary variables y_{inr}. Note that the variables y_{in} and y_{inr} are related as follows:

$$y_{inr} \leq y_{in}, \quad \forall i, n, r, \tag{17}$$

which means that alternative i is not available at scenario level if it is not made available by the operator or considered by the customer.

Discounted utility The basic behavioral assumption states that the customer selects the alternative associated with the largest utility. In order to avoid that an alternative that is not available is associated with the highest utility, we introduce the concept of *discounted utility*, that is equal to the utility when the alternative is available, and to a low value otherwise. The discounted utility associated by customer n with alternative i for scenario r is denoted by z_{inr} and defined as

$$z_{inr} = \begin{cases} U_{inr} & \text{if } y_{inr} = 1, \\ \ell_{nr} & \text{if } y_{inr} = 0, \end{cases} \quad \forall i, n, r, \tag{18}$$

where

$$\ell_{nr} = \min_{j \in \mathcal{C}} \ell_{jnr} \tag{19}$$

is the smallest lower bound across all alternatives. The linear formulation of (18) is given by

$$\ell_{nr} \leq z_{inr}, \qquad \forall i, n, r, \tag{20}$$

$$z_{inr} \leq \ell_{nr} + M_{inr} y_{inr}, \qquad \forall i, n, r, \tag{21}$$

$$U_{inr} - M_{inr}(1 - y_{inr}) \leq z_{inr}, \qquad \forall i, n, r, \tag{22}$$

$$z_{inr} \leq U_{inr}, \qquad \forall i, n, r, \tag{23}$$

where

$$M_{inr} = m_{inr} - \ell_{nr}. \tag{24}$$

To prove the equivalence between (18) and (20)–(23), we consider two cases:

- If $y_{inr} = 0$, constraints (20)–(23) become

$$\ell_{nr} \leq z_{inr}, \tag{25}$$

$$z_{inr} \leq \ell_{nr}, \tag{26}$$

$$U_{inr} - M_{inr} \leq z_{inr}, \tag{27}$$

$$z_{inr} \leq U_{inr}. \tag{28}$$

  Constraints (25) and (26) impose that $z_{inr} = \ell_{nr}$. Using (24), constraint (27) is written

$$U_{inr} - m_{inr} + \ell_{nr} \leq \ell_{nr}$$

  which is always verified from the definition (16) of the upper bound m_{inr}. Constraint (28) is written

$$\ell_{nr} \leq U_{inr},$$

  which is always verified as ℓ_{nr} is the smallest lower bound (see (16) and (19)).

- If $y_{inr} = 1$, constraints (20)–(23) become

$$\ell_{nr} \leq z_{inr}, \tag{29}$$

$$z_{inr} \leq \ell_{nr} + M_{inr}, \tag{30}$$

$$U_{inr} \leq z_{inr}, \tag{31}$$

$$z_{inr} \leq U_{inr}. \tag{32}$$

  Constraints (31) and (32) impose that $z_{inr} = U_{inr}$. Constraint (29) is written

$$\ell_{nr} \leq U_{inr},$$

  which is always verified (same argument as above). Using (24), constraint (30) is written

$$U_{inr} \leq \ell_{nr} + m_{inr} - \ell_{nr},$$

  which is always verified from the definition (16) of the upper bound m_{inr}.

Choice The choice of customer n in scenario r is characterized by the following binary variables:

$$w_{inr} = \begin{cases} 1 & \text{if } i \text{ is chosen,} \\ 0 & \text{otherwise,} \end{cases} \quad \forall i, n, r. \tag{33}$$

As each customer is choosing exactly one alternative, we impose

$$\sum_{i \in \mathcal{C}} w_{inr} = 1, \quad \forall n, r. \tag{34}$$

Since an alternative that is not available cannot be selected, we add the following constraint to the model:

$$w_{inr} \leq y_{inr}, \quad \forall i, n, r. \tag{35}$$

Based on the behavioral assumption, the chosen alternative corresponds to the one with the highest discounted utility. We introduce a continuous variable to capture it. It is denoted by U_{nr} and is defined as

$$U_{nr} = \max_{i \in \mathcal{C}} z_{inr}, \quad \forall n, r. \tag{36}$$

The linear formulation of (36) is given by

$$z_{inr} \leq U_{nr}, \qquad \forall i, n, r, \tag{37}$$

$$U_{nr} \leq z_{inr} + M_{nr}(1 - w_{inr}), \qquad \forall i, n, r, \tag{38}$$

where

$$M_{nr} = m_{nr} - \ell_{nr} \tag{39}$$

is the difference between the largest upper bound and the smallest lower bound, where the largest upper bound is defined as

$$m_{nr} = \max_{j \in \mathcal{C}} m_{jnr}. \tag{40}$$

To prove the equivalence between (33) and the formulation (37)–(38), we consider two cases:

- If $w_{inr} = 0$, constraints (37) and (38) become

$$z_{inr} \leq U_{nr}, \tag{41}$$

$$U_{nr} \leq z_{inr} + M_{nr}. \tag{42}$$

  Constraint (41) is consistent with the definition (36). Using (39), constraint (42) is written

$$U_{nr} \leq m_{nr} + (z_{inr} - \ell_{nr}),$$

  which is always verified as $z_{inr} - \ell_{nr} \geq 0$.

- If $w_{inr} = 1$, constraints (37) and (38) become

$$z_{inr} \leq U_{nr}, \quad (43)$$
$$U_{nr} \leq z_{inr}, \quad (44)$$

which implies that $U_{nr} = z_{inr}$, meaning that alternative i is associated with the highest discounted utility.

Expected demand The above formulation allows to represent the total expected demand of alternative $i \in \mathcal{C}$ by averaging over the number of considered scenarios:

$$D_i = \frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} w_{inr}. \quad (45)$$

Capacity allocation If the demand for alternative i is larger than its capacity, it is necessary to decide which customers have access to the alternative. In this framework, we have decided to model it exogenously, using an externally defined priority list of customers, similarly to Binder et al. (2017). A customer has access to an alternative if all customers before her in the list have also access to it. Note that the construction of this priority list can consider various aspects of the relationship between the operator and the customers, such as fidelity programs, VIP customers, etc. Therefore, the numbering of customers is important and reflects the priority list.

Based on the definition of the individual choice sets $\mathcal{C}_n$, we denote by $\mathcal{N}_i$ the set of customers considering alternative i, i.e., $\mathcal{N}_i = \{n \geq 1 | i \in \mathcal{C}_n\}$. By assuming that $\mathcal{N}_i$ is ordered according to the priority list, the constraint referring to the order of customers is written as

$$y_{inr} \leq y_{in^-r}, \quad \forall i, n \in \mathcal{N}_i, n > 1, r, \quad (46)$$

which says that if customer n^-, the predecessor of n in $\mathcal{N}_i$, does not have access to alternative i for scenario r, then neither does customer n.

The capacity restrictions are expressed by the following set of constraints:

$$\sum_{m \leq n^-} w_{imr} \leq (c_i - 1)y_{inr} + (n-1)(1 - y_{inr}), \qquad \forall i, n \in \mathcal{N}_i, n > c_i, r, \quad (47)$$
$$c_i(y_{in} - y_{inr}) \leq \sum_{m \leq n^-} w_{imr}, \qquad \forall i, n \in \mathcal{N}_i, n > 1, r. \quad (48)$$

Constraint (47) ensures that the capacity cannot be exceeded, and it can be verified by considering two cases:

- If $y_{inr} = 0$, (47) is written

$$\sum_{m \leq n^-} w_{imr} \leq n - 1,$$

which is always satisfied.

- If $y_{inr} = 1$, (47) is written

$$1 + \sum_{m \leq n^-} w_{imr} \leq c_i,$$

which means that the number of customers up to and including n who have chosen alternative i does not exceed c_i, so there is still room for customer n.

Constraint (48) forbids the access of customers to a certain alternative when its capacity has been reached. To verify the validity of this constraint we consider the three possible cases for the values of y_{inr} and y_{in} (note that $y_{inr} = 1$ and $y_{in} = 0$ is infeasible due to constraint (17)):

- If $y_{inr} = 1$, then $y_{in} = 1$ (because of (17)), and (48) is written

$$\sum_{m \leq n^-} w_{imr} \geq 0, \tag{49}$$

which is clearly satisfied.

- If $y_{inr} = 0$ and $y_{in} = 0$, (48) is also written as (49).
- If $y_{inr} = 0$ and $y_{in} = 1$, (48) is written

$$\sum_{m \leq n^-} w_{imr} \geq c_i,$$

which implies that the capacity has been reached due to the choices of customers in $\mathcal{N}_i$ up to and including n^-, and even if the alternative is proposed to customer n by the operator ($y_{in} = 1$), there is no room left for her.

Note that if alternative i is not offered by the operator ($y_i = 0$), the variables y_{in}, y_{inr} and w_{inr} are equal to 0 $\forall n \in \mathcal{N}_i$ (due to constraints (5), (17) and (35), respectively), and therefore constraints (46)–(48) are always satisfied.

3.4 Capacities as decision variables

In terms of capacity allocation, although capacities c_i are supposed to be given, the formulation can easily be extended to include capacity as a decision variable. In order to avoid the nonlinearity that would appear in the capacity constraints (47) and (48), the model should be specified as follows. For each decision variable c_i, a predefined list of Q feasible values for the capacity must be proposed: $c_{i1}, \ldots, c_{iQ}$. Then, alternative i is duplicated Q times, each instance being associated with the same utility function, but with a different capacity. It is sufficient to include the constraint

$$\sum_{q=1}^{Q} y_{iq} \leq 1, \quad \forall i, \tag{50}$$

to guarantee that at most one of the duplicates is actually available. The variables y_{iq} are the extension of the variables y_i, and control the actual capacity of the facility. Note that it is still possible for the operator to decide not to open it. In that case, the sum on the left hand side of (50) would be zero.

The rest of the variables and constraints introduced so far have to be adapted accordingly. The binary variables y_{in}, y_{inr}, z_{inr} and w_{inr} are replaced by the variables y_{inq}, y_{inrq}, z_{inrq} and w_{inrq}, respectively. Constraints (17), (20)–(23), (35), (37)–(38) and (46)–(48) need to incorporate the index q. Constraint (34) is written

$$\sum_{i \in \mathcal{C}} \sum_{q=1}^{Q} w_{inrq} = 1, \quad \forall n, r, \tag{51}$$

and the expected demand of alternative i is written

$$D_i = \frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{q=1}^{Q} w_{inrq}. \tag{52}$$

Since the utility function is the same, both the utility variables U_{inr} and the associated constraint (15) remain unchanged.

The number of constraints involved in this specification is of the order of $JNRQ$, where $J = |\mathcal{C}|$ is the number of alternatives within the choice set $\mathcal{C}$. In real applications, where the number of customers can be really large, this comes with a high computational price. In order to reduce the size of the model, customers can be grouped into classes of homogeneous behavior. A synthetic population, which is constructed by combining different data sources, is convenient here (Farooq et al.; 2013).

The complexity of the probability distributions of the random variables involved in the choice models and their correlation structure are irrelevant in this context as long as it is possible to draw from these distributions. Indeed, the generation of draws is performed at a preprocessing stage. The complexity of the formulation is only affected by the number of draws, and not by the nature of the underlying distributions. This is a strength of the framework, that is relevant for any existing complex model, and for other models to be developed in the future.

The supply, in terms of the alternatives offered to customers, is flexible in the sense that there exists the possibility not to propose an alternative i thanks to the variables y_i. This allows the operator to investigate marketing solutions and business models. In certain contexts, it might be interesting to extend this feature in order to be able to propose an alternative to some customers but not to others. This will increase the complexity of the model, especially when it comes to identify consecutive customers who are competing for the same alternative, and it is out of the scope of the paper.

The formulation developed in this section is linear, and applies to a great deal of choice models. It can be integrated in any MILP model derived from a relevant OR application. To exemplify that, we consider the example of a benefit maximization problem in Section 4.

4 Demand-based benefit maximization

A relevant application where the use of advanced demand models plays an important role is the maximization of benefit, understood as the difference between the generated revenue and the operating costs. We use this approach to illustrate how the framework described in Section 3 can be employed. Concrete examples can be found in transportation, for instance, where services are offered in a competitive market. In such cases, it is expected that a detailed representation of the heterogeneity of the demand might lead to a better planning of the associated transport operations.

We are aiming at finding the best strategy in terms of pricing and capacity allocation in order to maximize the benefit of the operator. We assume that it is selling services to a market, each of them at a certain capacity and at a certain price, both to be decided. With respect to the cost of each service, we assume that it is composed of a fixed cost associated with operating the service and a variable cost associated with each sold unit of the service.

The market is composed of N customers, which are assumed to be heterogeneous and price elastic, in the sense that each costumer may have a different behavior and sensitivity towards price. The operator is considering J services to offer, each of them having Q associated levels of capacity, as described in Section 3.4.

In a benefit maximization context, we need to model competition. Indeed, if we do not account for competitive alternatives, customers are captive, and the benefit maximization problem is unbounded. Competitive alternatives can be explicitly modeled in the choice set, or grouped into an *opt-out* alternative that captures customers leaving the market, either because they choose an alternative from a competitor or because they do not choose anything at all. Here we consider the second approach. The main assumption behind is that the decisions of the competitors are given, and not adjusted as a consequence of the decisions of the operator. The opt-out option is denoted by $i = 0$. Note that it is always available to all customers, i.e., $0 \in \mathcal{C}_n \; \forall n \geq 1$.

We consider the price as an endogenous variable in the utility function (15). We define $p_{in} \in \mathbb{R}$ as the price that customer n must pay to access service $i \in \mathcal{C}_n \setminus \{0\}$. Note that the index n allows the operator to propose different prices to different customers or, more realistically, to different groups of customers (e.g., students, seniors, families). In that case, the model includes as many p variables as the number of groups.

The expected revenue obtained from service $i > 0$ can be derived directly from the demand expression (52) and the price specification:

$$R_i = \frac{1}{R} \sum_{n \in \mathcal{N}_i} \sum_{r=1}^{R} \sum_{q=1}^{Q} p_{in} w_{inrq}. \tag{53}$$

As the price is an endogenous variable, (53) is nonlinear. The product of a binary and a continuous variable can be easily linearized if an upper bound for the continuous variable is known, which in this case can be assumed by the operator. However, it is in general more convenient, both from an application and a modeling point of view, to assume that p_{in} can only take a finite number of predetermined values, called price levels.

Note that any integer variable that is bounded and can take only a finite number of values can be written as a linear combination of binary variables. In our case, p_{in} is not defined as an integer variable, and neither its price levels, but they can be expressed as integer numbers by setting a precision of k decimals and multiplying them by 10^k.

Consider $p_{in} \in 1/10^k\{\ell_{in}, \dots, m_{in}\}$, where $\{\ell_{in}, \dots, m_{in}\}$ are the integer price levels for customer n and service $i \in \mathcal{C}_n \setminus \{0\}$, sorted from the smallest level (ℓ_{in}) to the largest (m_{in}). We define L_{in}

binary variables $\lambda_{in\ell}$ for each customer n and service $i \in \mathcal{C}_n \setminus \{0\}$, where L_{in} is the smallest integer such that $m_{in} - \ell_{in} \leq 2^{L_{in}} - 1$, that is

$$L_{in} = \lceil \log_2(m_{in} - \ell_{in} + 1) \rceil. \tag{54}$$

We can then write p_{in} as follows:

$$p_{in} = \frac{1}{10^k}\left(\ell_{in} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell}\right). \tag{55}$$

If it is important to generate prices below m_{in}, the following constraint must be included:

$$\ell_{in} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell} \leq m_{in}, \quad \forall n, i \in \mathcal{C}_n \setminus \{0\}. \tag{56}$$

The expected revenue R_i is now written as

$$R_i = \frac{1}{R}\frac{1}{10^k} \sum_{n \in \mathcal{N}_i} \sum_{r=1}^{R} \sum_{q=1}^{Q} \left(\ell_{in} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell}\right) w_{inrq}. \tag{57}$$

In order to linearize the product of the binary variables $\lambda_{in\ell}$ and w_{inrq}, we introduce the binary variables $\alpha_{inrq\ell} = \lambda_{in\ell} w_{inrq}$, so that R_i becomes linear:

$$R_i = \frac{1}{R}\frac{1}{10^k}\left[\sum_{n \in \mathcal{N}_i} \sum_{r=1}^{R} \sum_{q=1}^{Q} \left(\ell_{in} w_{inrq} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \alpha_{inrq\ell}\right)\right], \tag{58}$$

with the linearizing constraints

$$\lambda_{in\ell} + w_{inrq} \leq 1 + \alpha_{inrq\ell}, \qquad \forall n, i \in \mathcal{C}_n \setminus \{0\}, r, q, \ell, \tag{59}$$

$$\alpha_{inrq\ell} \leq \lambda_{in\ell}, \qquad \forall n, i \in \mathcal{C}_n \setminus \{0\}, r, q, \ell, \tag{60}$$

$$\alpha_{inrq\ell} \leq w_{inrq}, \qquad \forall n, i \in \mathcal{C}_n \setminus \{0\}, r, q, \ell. \tag{61}$$

Regarding the costs, we assume that the operating cost of service $i > 0$ is calculated as

$$C_i = \sum_{q=1}^{Q} (f_{iq} + v_{iq} c_{iq}) y_{iq}, \tag{62}$$

where f_{iq} is the fixed cost and v_{iq} is the cost per sold unit of service i with capacity level c_{iq}.

The total benefit is computed by subtracting from the generated revenues the total operating costs. These quantities are obtained by adding the revenues and costs from all the services proposed by the operator (given by (58) and (62), respectively) except the opt-out option (i.e., the services proposed by the competitors are excluded):

$$\max \sum_{i>0} (R_i - C_i) = \sum_{i>0} \left(\frac{1}{R}\frac{1}{10^k}\left[\sum_{n \in \mathcal{N}_i} \sum_{r=1}^{R} \sum_{q=1}^{Q} \left(\ell_{in} w_{inrq} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \alpha_{inrq\ell}\right)\right] - \sum_{q=1}^{Q} (f_{iq} + v_{iq} c_{iq}) y_{iq} \right). \tag{63}$$

The constraints of the model are itemized next:

- Utility: (15),
- Availability: (5), (17),
- Discounted utility: (20), (21), (22), (23),
- Choice: (35), (37), (38), (51),
- Capacity allocation: (46), (47), (48), (50) and
- Pricing: (56), (59), (60), (61).

Note that the size of the above optimization model can be large (see final discussion in Section 3). Nevertheless, the structure of the model is particularly well suited for decomposition methods. Indeed, most constraints are independent across customers, and across scenarios. The coupling occurs only when the benefit is calculated. Investigating decomposition methods is out of the scope of the paper, and left for future research.

In Section 5, we illustrate the use of the model with a case study of moderate size, based on a real choice model taken from the recent literature.

5 Case study

We consider the case study of a parking services operator, which is motivated by the availability of a published disaggregate choice model for parking choice (Ibeas et al.; 2014). For the sake of illustration, and to avoid solving huge optimization problems, the size of the sample under consideration is $N = 50$ customers. The choice set consists of three services: paid on-street parking (PSP), paid parking in an underground car park (PUP) and free on-street parking (FSP). Since the latter does not provide any revenue to the operator, it is considered as the opt-out option.

The objective of the operator is to maximize the benefit obtained from PSP and PUP. Together with the technical variables used for the linearization of the constraints, the decision variables of the model for this case study are characterized as follows:

- Availability variables y_{iq}: We assume $y^d_{in} = 1\ \forall i, n$, that is, all customers consider all parking facilities when deciding where to park. Therefore, $y_{inq} = y_{iq}\ \forall n$ (see (5));
- Availability variables y_{inrq};
- Choice variables w_{inrq};
- Price variables $\lambda_{in\ell}$: We do not offer a different price for each customer (p_{in}), but a general price for everyone (p_i). Therefore, the price variables to be considered are $\lambda_{i\ell}$. Furthermore, we assume that the number of price levels is the same for both PSP and PUP ($L_{PSP} = L_{PUP} = L$).

The demand model is the parking choice model proposed in Ibeas et al. (2014), whose data was kindly provided by the authors and used to perform the experiments discussed in the upcoming sections. They characterize a mixture of logit models to describe the behavior of potential car park users when choosing a parking place.

The specification table of the utility functions is provided in Table 1. The random coefficients are those associated with the access time to the parking place once the user arrives to the parking area (AT) and the parking fee (FEE). As mentioned by the authors, the latter is related to an hour of use of the parking place, regardless of the time that the spot was needed. The units are not specified. For these two variables the associated parameters β_{AT} and β_{FEE} are assumed to be normally distributed. Furthermore, these coefficients are correlated, with $cov(\mathsf{AT}, \mathsf{FEE}) = -12.8$.

		FSP	PSP	PUP
ASC_{PSP}	32	0	1	0
ASC_{PUP}	34	0	0	1
β_{AT}	$\sim N(-0.788, 1.06)$	AT_{FSP}	AT_{PSP}	AT_{PUP}
β_{TD}	-0.612	TD_{FSP}	TD_{PSP}	TD_{PUP}
$\beta_{Origin_{INT\_FSP}}$	-5.76	$Origin_{INT\_FSP}$	0	0
β_{FEE}	$\sim N(-32.3, 14.2)$	0	FEE_{PSP}	FEE_{PUP}
$\beta_{FEE_{PSP(LowInc)}}$	-11	0	$FEE_{PSP}LowInc$	0
$\beta_{FEE_{PSP(Resident)}}$	-11.4	0	$FEE_{PSP}Resident$	0
$\beta_{FEE_{PUP(LowInc)}}$	-13.7	0	0	$FEE_{PUP}LowInc$
$\beta_{FEE_{PUP(Resident)}}$	-10.7	0	0	$FEE_{PUP}Resident$
$\beta_{AgeVeh\leq 3}$	4.04	0	0	$AgeVeh \leq 3$

Table 1: Specification table of the mixed logit model

The other variables appearing in the utility specification are the following: access time to the destination from the parking spot (TD), a dummy variable that is 1 if the origin of the trip is internal to the town ($Origin_{INT\_FSP}$), a dummy variable that is 1 if the income of the customer is below 1200€/month ($LowInc$), a dummy variable that is 1 if the customer is resident ($Resident$) and a dummy variable that is 1 if the age of the vehicle is lower than 3 years ($AgeVeh_{\leq 3}$). Two interactions to address the variations in taste among customers are also considered: FEE with having a low income and FEE with being resident. Note that in the optimization model FEE is the only endogenous variable, and all others are exogenous demand variables.

This section is divided into four experiments. In the first three, we assume that the paid services have a fixed capacity of $c_{PSP} = c_{PUP} = 20$ spots each. This value is large enough so that it is realistic for the size of the sample but restrictive enough so that some customers are forced to opt-out because there is not enough room for everyone. Therefore, the index q is dropped from the formulation and both the fixed and variable costs are equal to 0, which converts the objective function into revenue maximization. This will decrease the complexity of the model for the first tests. More precisely, Section 5.1 characterizes the optimal prices of the parking services as well as their expected demands, Section 5.2 describes two scenarios for price differentiation by market segmentation, and Section 5.3 evaluates the impact of the priority list on the optimal prices and expected revenue. The last experiment is presented in Section 5.4, and analyzes the effects of considering the cost of capacity allocation. Here we assume 4 different capacity levels for both PSP and PUP. In all cases, FSP is assumed to have unlimited capacity and the priority list is defined as the order of the customers in the original dataset, which can be interpreted as a random arrival.

The computer codes used in our experiments have been implemented in C++ with CPLEX 12.7 callable libraries. All experiments were performed using 6 threads in a 3.33 GHz Intel Xeon X5680 server running a 64-bit Ubuntu 16.04.2.

5.1 Price calibration

In this section, we aim at deciding the optimal price for both parking services so that the revenue of the operator is maximized. We start by considering an unlimited capacity for the parking services, i.e., constraints (46)–(48) are ignored, and afterwards, we include them back in order to analyze the expected increase of solution time due to the increase in complexity, and the changes in the obtained results.

We determine 16 price levels ($L = 4$) for PSP and PUP based on the values of the variable FEE in the dataset (0.6 and 0.8 respectively):

- PSP: $\{0.50, 0.51, \dots, 0.65\}$, and
- PUP: $\{0.70, 0.71, \dots, 0.85\}$.

Note that these levels are defined in a way that all combinations of the associated binary variables are feasible (i.e., constraint (56) is not needed). Indeed, there is no reason to discard any price level because we do not have any external restriction on the price.

Figure 1 shows the variation of solution time (in a logarithmic scale) and expected revenue with respect to the number of draws R, and Figure 2 shows the variation of price and expected demand of each service, both for the uncapacitated case. In general terms, the values for the demand and the total revenue stabilize as the number of scenarios increases. In the case of PSP, the price is set to 0.54 in the last 4 instances, and for PUP it is set to 0.74 in the last two. However, the increase in solution time is considerable, as expected. For example, from $R = 100$ to $R = 250$ it changes from 21 minutes to almost 2.5 hours.

For the capacitated case, Figure 3 shows that the increase in solution time is high compared to the uncapacitated case for the large instances. For instance, for $R = 100$ the solution time increases from 21 minutes to more than 6 hours, and for $R = 250$ from 2.5 hours to almost 2 days. This manifests that the implementation of the priority list and the tracking of the occupancy for each alternative hugely complicate the solution approach.

Regarding the optimal prices, we can see in Figure 4 that both PSP and PUP are more expensive in the capacitated case. Since the demand of PSP was already higher in the uncapacitated case than its current capacity, its price can be increased so that the operator obtains a higher revenue from the customers accessing the service. In the case of PUP, the price is also higher, but the demand is similar to the one in the uncapacitated case, which might also be influenced by the capacity restriction on PSP, since normally the opt-out option is the least attractive one. This service is experiencing an increase in the demand because it is capturing the customers that cannot be allocated due to capacity limitations or that are not willing to pay the current price of the paid alternatives.

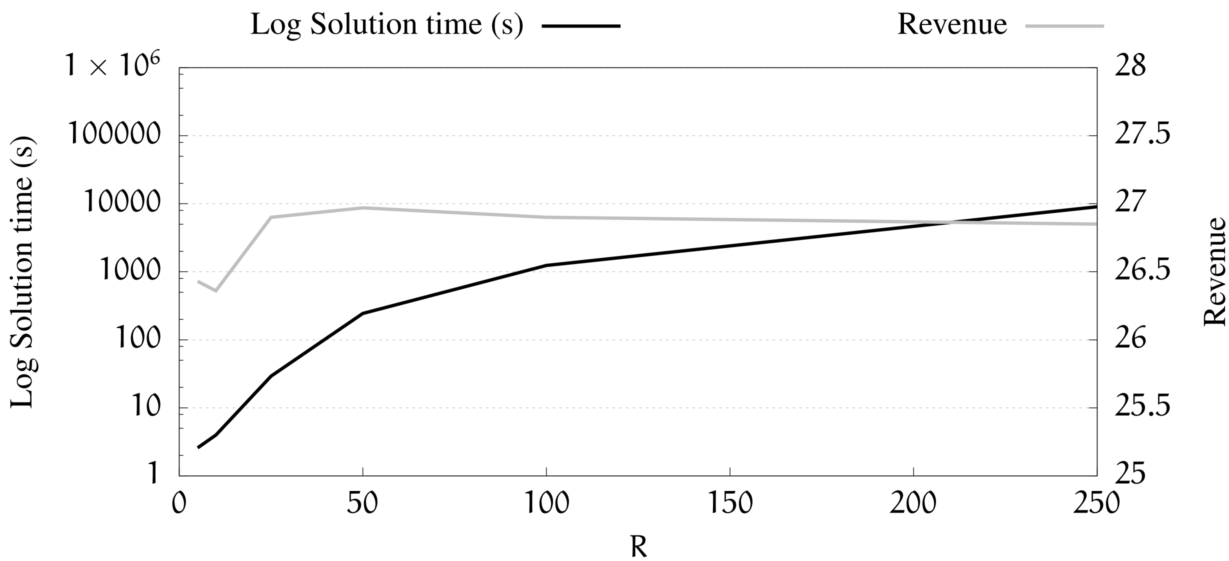

Figure 1: Variation of solution time and revenue with respect to the number of draws for the uncapacitated case

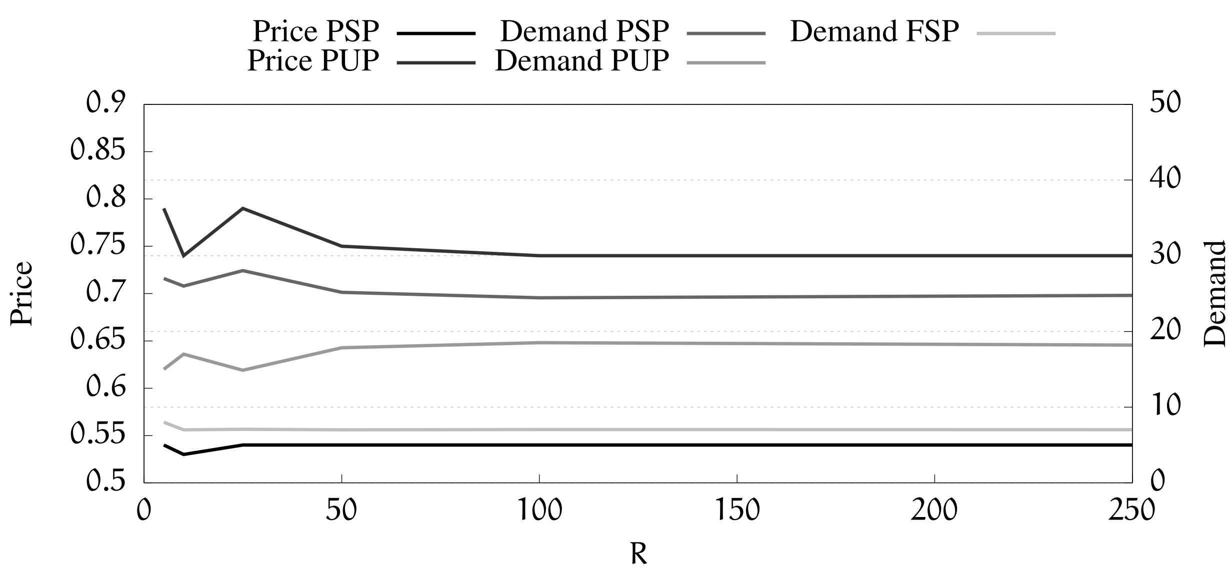

Figure 2: Variation of price and demand with respect to the number of draws for the uncapacitated case

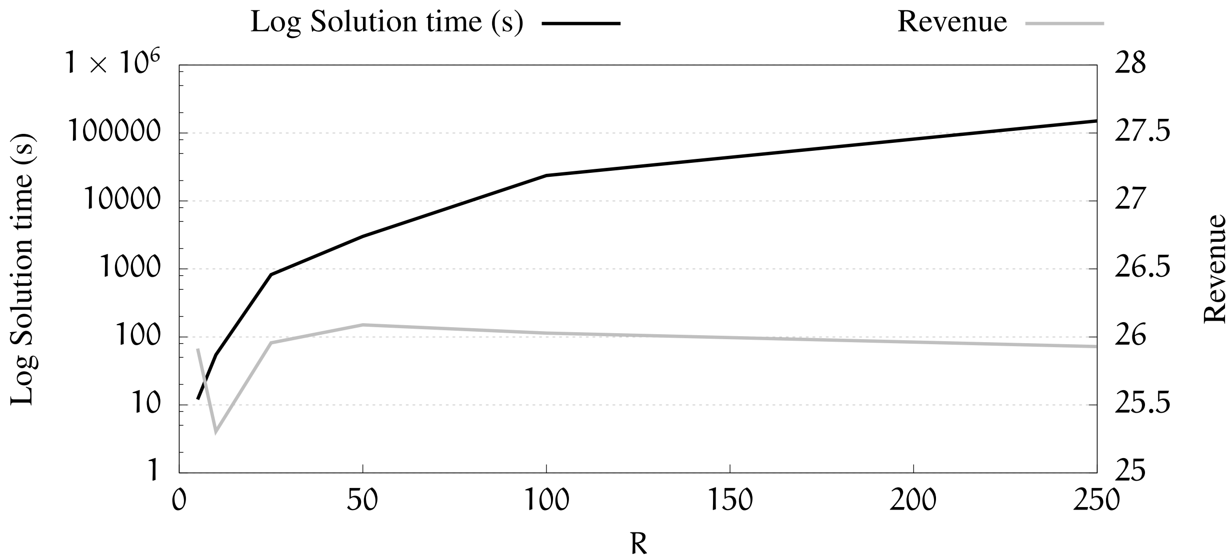

Figure 3: Variation of solution time and revenue with respect to the number of draws for the capacitated case

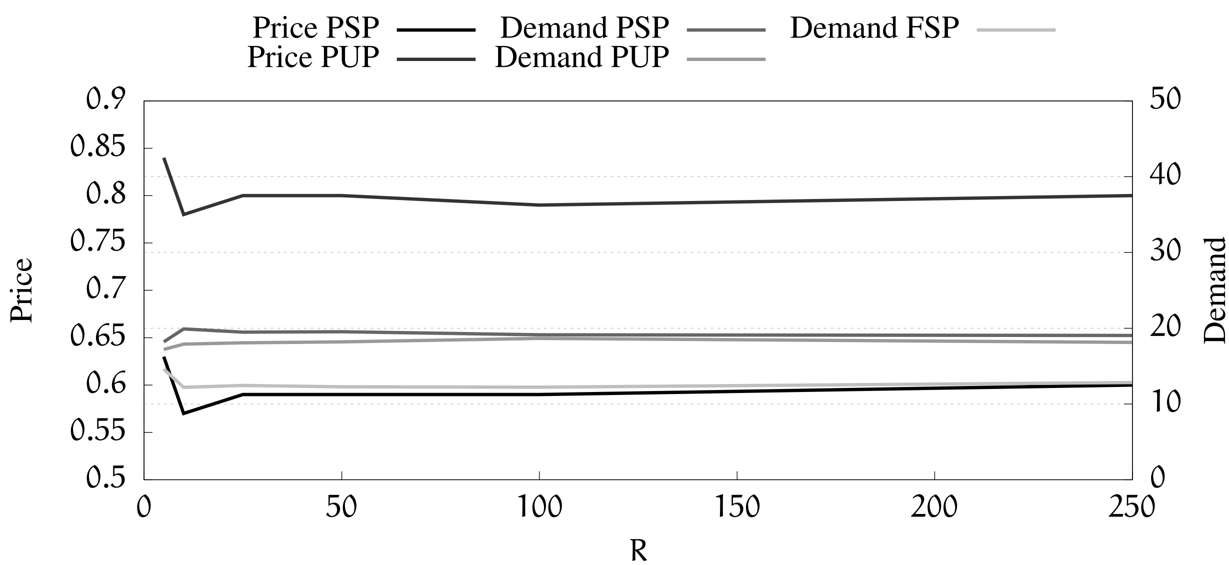

Figure 4: Variation of price and demand with respect to the number of draws for the capacitated case

The values of the solution time, optimal prices, expected demand and expected revenue for the described instances can be found in Appendix A.1 (see Tables 3 and 4). Since the results stabilize after 50 draws, the experiments performed in the remainder of the paper consider $R = 50$.

5.2 Price differentiation by population segmentation

Imagine now that the municipality provides reduced fees to residents (R) who want to access one of the paid alternatives. This is actually done in many cities, where residents get reduced prices for common parking services or even exclusive areas, where only them have the right to park. In this case, we assume a discount factor that is applied to the prices offered to non residents (NR). Regarding the revenue obtained by the operator, two scenarios are considered:

1. The difference between the actual price of the service and the contribution of the resident is paid by the municipality in the form of a subsidy and, therefore, contributes to the revenue of the operator.

2. The operator is obliged by the municipality to offer reduced fees to residents, without any other compensation than the right to operate the parking.

Note that in the first scenario, the reduced prices have an impact on the utility functions of the residents, and therefore on their choice, but not on the revenue of the operator. In the second case, the reduced fares cause a decrease in the total revenue.

Since residents only pay a part of the fee that non residents pay, the former customers might be attracted to higher fares, and we expect the prices of both services to increase. Therefore, we modify the price levels for both PSP and PUP accordingly:

- PSP: $\{0.60, 0.64, \ldots, 1.20\}$, and
- PUP: $\{0.80, 0.84, \ldots, 1.40\}$.

Note that we keep the same number of price levels ($L = 16$), but we change the step between price levels from 0.1 to 0.4 to extend the range.

The optimal prices for different discount rates are included in Figures 5 and 6 for scenarios 1 and 2, respectively. In both cases, the higher the discount, the higher the prices offered to non residents, as expected. However, in the second scenario, this increase in the price is more moderate because it leads to a decrease in the total revenue. We note that when the discount substantially increases, the prices in scenario 1 are larger than those in scenario 2, as well as the number of residents choosing the paid alternatives, as shown in Figures 7 and 8.

In terms of demand, we see that the higher the discount, the higher the number of non residents deciding to opt-out, because they are not willing to pay the offered fares and they choose FSP instead. We also note that the demand values for PSP and PUP remain similar among instances in both scenarios (for PSP they oscillate between 18.5 and 19.5, and for PUP between 17.9 and 18.9, with an exception of 15.7, see Tables 5 and 6 in Appendix A.2), but with increasing number of residents and decreasing number of non residents with the increase in the discount.

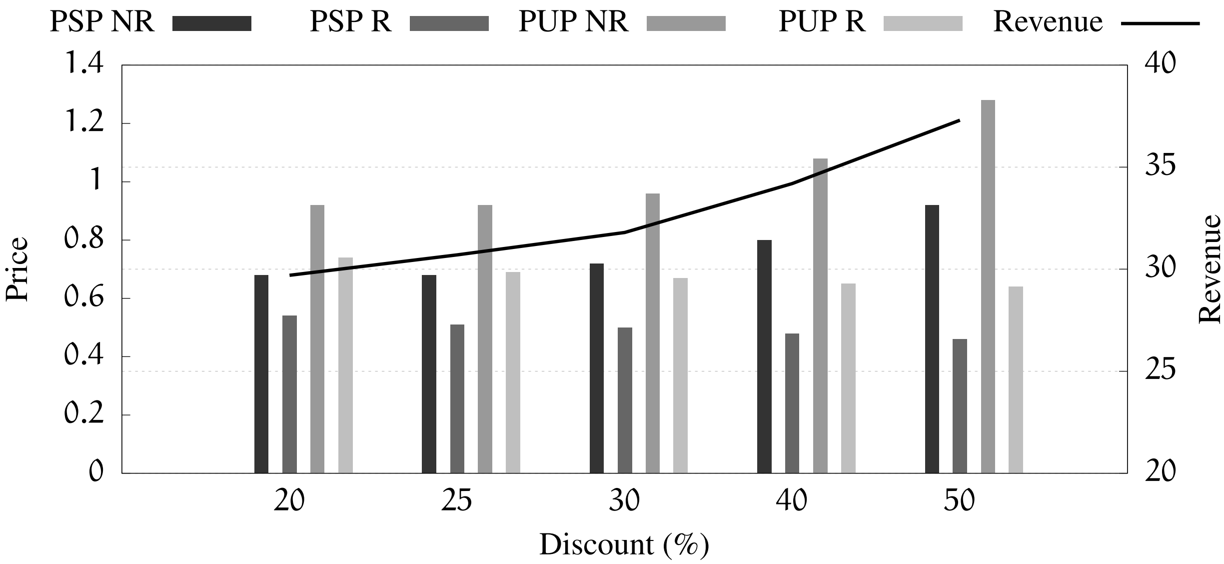

Figure 5: Prices per segment for different discount rates in scenario 1

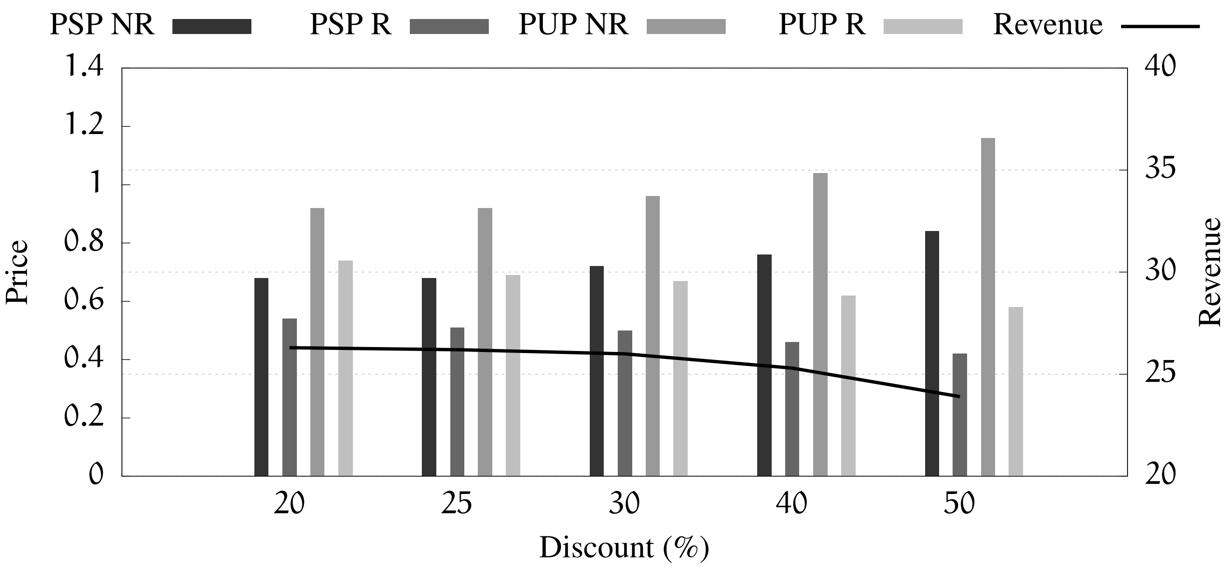

Figure 6: Prices per segment for different discount rates in scenario 2

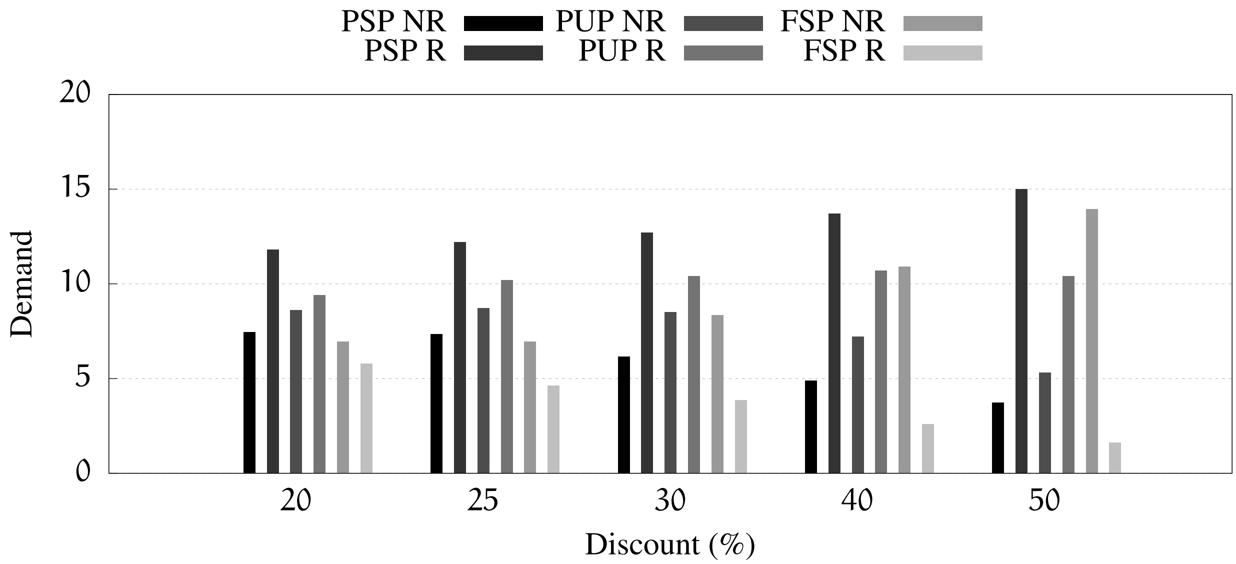

Figure 7: Demand per segment and revenue for different discount rates in scenario 1

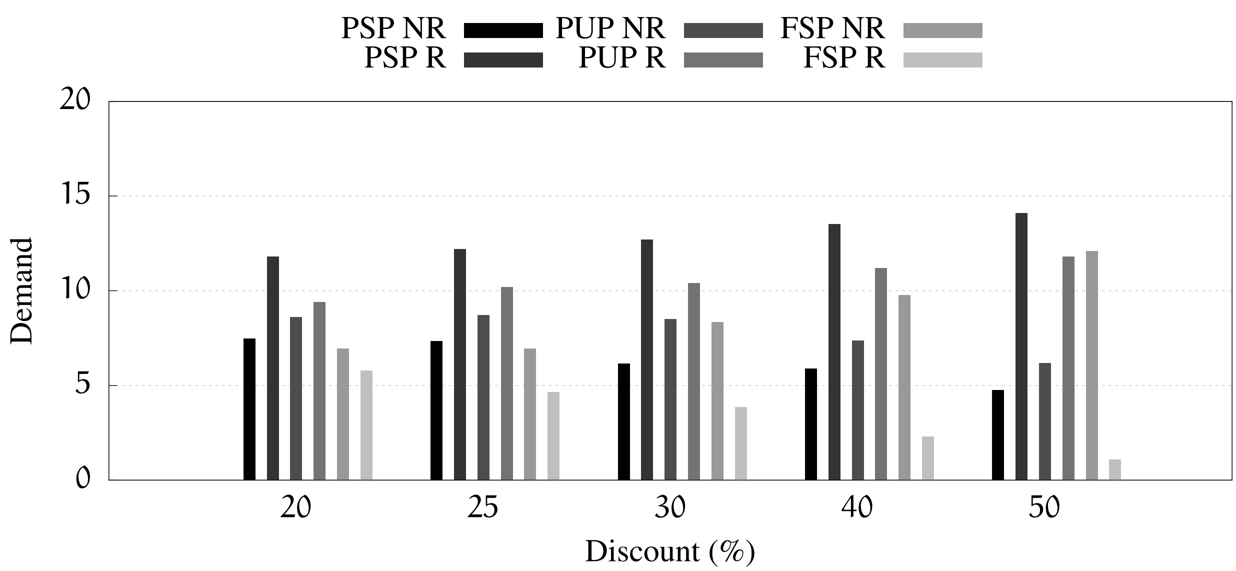

Figure 8: Demand per segment and revenue for different discount rates in scenario 2

5.3 Impact of the arrival of customers

The priority list described in Section 3.3 states the order in which customers are considered to access the services. As mentioned above, the priority list in this case study is defined as the order of customers in the data. We can analyze the impact of such priority list on the obtained results by testing different priority lists. For this experiment we consider the second scenario in Section 5.2, i.e., the operator is forced by law to offer a discount on the fares for residents. For the sake of simplicity we only consider the 30% discount rate.

We construct 100 different priority lists by shuffling the customers, so that they might arrive in a different order. We also run these instances with $R = 50$. Figures 9 and 10 show the distribution of the prices for the paid alternatives and the expected revenue, respectively.

The results show that the framework performs similarly in terms of the obtained results when a random arrival of the customers is assumed. Regarding the prices, the same price levels are determined for almost all the instances for both PSP and PUP, and with respect to the obtained revenues, the variation is small, with values that oscillate between 25.7 and 26.1. This is consistent with the findings of Binder et al. (2017), who show that the aggregate indicators are stable across realizations of a random priority list.

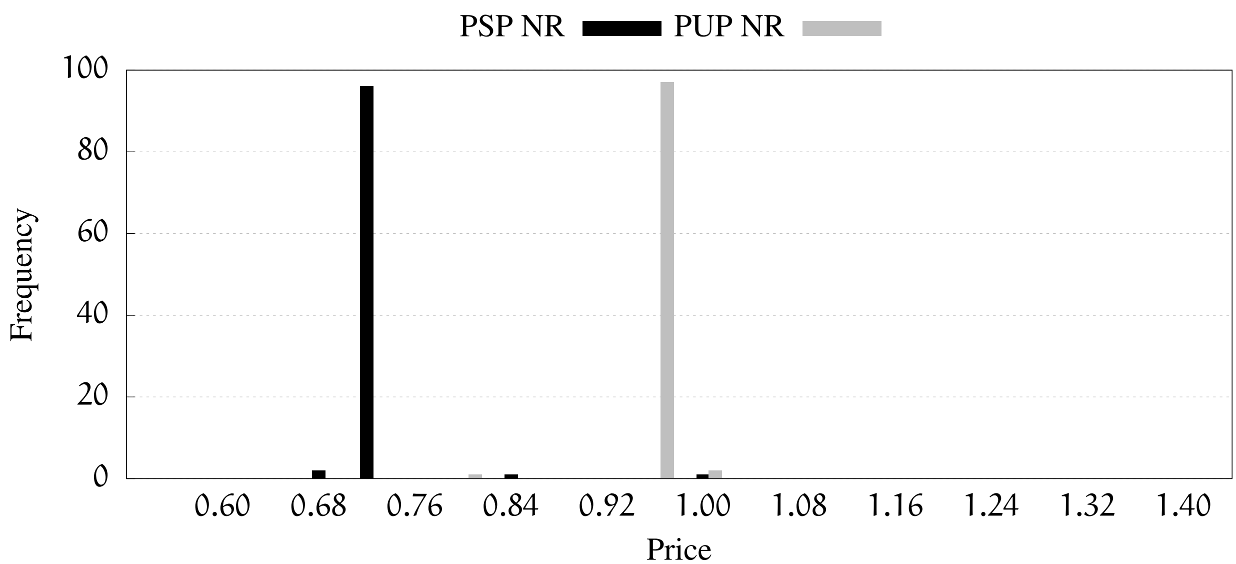

Figure 9: Histogram of prices offered to NR in scenario 2 for a discount rate of 30%

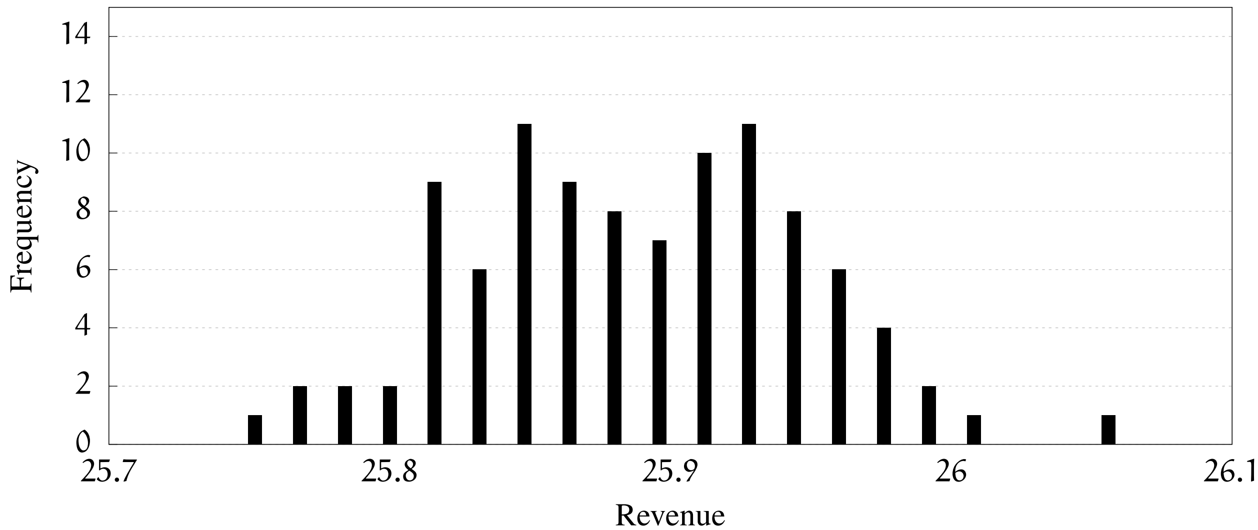

Figure 10: Histogram of revenues in scenario 2 for a discount rate of 30%

We have also tested priority lists set deterministically. We have considered two extreme cases, one in which all residents have priority over the non residents and the other way around, which can be interpreted as the arrival of all residents first and last, respectively. A higher revenue is expected when the residents arrive last, since they produce a lower revenue (because of the offered discount) and they leave room for non residents. For this case study, the obtained results are very similar, with the revenue gained from residents coming last slightly higher.

5.4 Benefit maximization through capacity allocation

In this experiment, we test 4 different capacity levels for both services. As described in Section 3.4, we replicate the services PSP and PUP as many times as capacity levels we want to evaluate. We consider 5, 10, 15 and 20 parking spots, which makes 4 copies of PSP and 4 copies of PUP, each of them with the same utility function but a different level of capacity. Together with the opt-out option (FSP), this experiment contains 9 different services.

Note that constraint (50) does not force the opening of both services, since it might be more convenient from a benefit maximization point of view to allocate all the resources of the operator to only one of the facilities. However, in case we want to make sure that both PSP and PUP are offered, we can modify this constraint as follows:

$$\sum_{q=1}^{Q} y_{iq} = 1, \quad \forall i > 0. \tag{64}$$

As mentioned in Section 4, the cost associated with operating a parking facility is composed of a fixed cost and a variable cost (in this case, a cost per parking spot). For the sake of illustration, we assume that both types of cost are the same among capacity levels, i.e.,

- PSP: $f_{PSP,q} = 1.5$ and $\nu_{PSP,q} = 0.35 \quad \forall q = 1, \ldots, 4$, and
- PUP: $f_{PUP,q} = 3$ and $\nu_{PUP,q} = 0.5 \quad \forall q = 1, \ldots, 4$.

The results for both approaches (constraint (50) and (64)) are included in Table 2. We see that in this case it makes sense to close PUP and only open PSP with the highest level of capacity. Indeed, when we impose that both parking facilities have to be opened, PUP is offered at the lowest capacity level available, and a lower benefit is generated. The obtained solution time gives us an idea of the increase in complexity with respect to the revenue maximization problem. Indeed, for $R = 50$, it goes from 50 minutes (capacitated revenue maximization) to almost 19 hours (benefit maximization with constraint (50)).

	Solution	**Capacity**		**Demand**			**Prices**		
Constraint	time (h)	PSP	PUP	PSP	PUP	FSP	PSP	PUP	Benefit
(50)	18.7	20	-	19.4	-	30.6	0.76	-	6.27
(64)	33.7	15	5	14.8	4.56	30.7	0.76	1.32	4.99

Table 2: Capacity and pricing characterization for approaches (50) and (64)

The experiments performed in Section 5 illustrate the great flexibility of the framework proposed in this paper. It has been employed to calibrate the prices of the services being offered, to offer different prices to different segments of the population, and to evaluate different capacity levels from a benefit maximization point of view. Indeed, the possibilities are endless, and features such as complex behavioral patterns and alternative supply configurations can be investigated with the framework.

6 Conclusions and future work

We have proposed a general formulation of an advanced choice model that is designed to be included in MILP. It is general in the sense that *any* assumption can be made on the probability distribution of the error term of the utility function, so that this approach is not limited to simple discrete choice models, and that it can be included in any optimization problem that requires a demand representation. The stochasticity of the model is captured by drawing from the distribution of the random variables involved in the model specification. This allows us to avoid the explicit formulation of choice probabilities and to work directly with the utility functions, using the first principles of utility maximization.

Regarding the supply aspects of the framework, an illustrative MILP model is developed in order to define a pricing strategy that maximizes the total benefit of the operator when the offered services have a certain capacity. The model takes into account the preferences of customers by means of the utility function when determining the prices of the services. Since the customers are not captive, i.e., they can leave the market by choosing the opt-out option, there exists a tradeoff between price and customers' choices.

The results exhibit that this formulation is a powerful tool to configure systems based on the heterogeneous behavior of customers. Some properties of the systems, such as the price, can be set specifically for different market segments, which tailors the systems to the users at the same time that the benefit is maximized.

Nevertheless, the disaggregate representation of customers' preferences, together with the linearity of the formulation, implies that the dimension of the resulting problem is high, and therefore solving it is computationally expensive. This is an issue that needs to be addressed because in practice, populations are large and a high number of draws is desirable to be as close as possible to the true value.

Decomposition techniques are convenient in this case in order to speed up the solution approach. The model, by design, can naturally be decomposed along two dimensions. On the one hand, each draw r constitutes an independent scenario, and all scenarios are considered together solely in the objective function (for the calculation of revenue). On the other hand, each customer n has an associated optimization problem, which consists of choosing the alternative among the available ones maximizing her utility. All customers are coupled in the capacity constraints to ensure that the capacity of each alternative is not exceeded, and in the objective function, to calculate the total demand.

To sum up, in this paper, we highlight the importance and significance of taking supply and demand interactions into account in an optimization framework from both operators' (benefit) and consumers' (convenience) points of view. The framework can be used in various contexts, such as facility location, revenue management, transportation, supply chain management, and logistics. The main challenge that researchers in the above mentioned contexts face while integrating choice models inside optimization problems concerns the definition of a closed-form tractable mathematical model. Here, we propose a state-of-the-art mathematical model that addresses this issue profoundly and sheds light for future research avenues in these topics.

Acknowledgements

This research is supported by the Swiss National Science Foundation Grant 200021-165636 "Incorporating advanced behavioral models in mixed integer linear optimization." The authors would like to thank the authors in Ibeas et al. (2014), who provided the dataset employed in the performed experiments.

7 Bibliography

Andersson, S.-E. (1998). Passenger choice analysis for seat capacity control: a pilot project in scandinavian airlines, *International Transactions in Operational Research* **5**(6): 471 – 486.

Aros-Vera, F., Marianov, V. and Mitchell, J. E. (2013). p-hub approach for the optimal park-and-ride facility location problem, *European Journal of Operational Research* **226**(2): 277–285.

Benati, S. (1999). The maximum capture problem with heterogeneous customers, *Computers & Operations Research* **26**(14): 1351–1367.

Benati, S. and Hansen, P. (2002). The maximum capture problem with random utilities: Problem formulation and algorithms, *European Journal of Operational Research* **143**(3): 518–530.

Binder, S., Maknoon, Y. and Bierlaire, M. (2017). Introducing exogenous priority rules for the capacitated passenger assignment problem, *Technical Report TRANSP-OR 170209*, Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland.

Bitran, G. and Caldentey, R. (2003). An overview of pricing models for revenue management, *Manufacturing & Service Operations Management* **5**(3): 203–229.

Farooq, B., Bierlaire, M., Hurtubia, R. and Flötteröd, G. (2013). Simulation based population synthesis, *Transportation Research Part B: Methodological* **58**(C): 243–263.

Francis, R. L. and Lowe, T. J. (2015). *Demand Point Aggregation for Some Basic Location Models*, Springer International Publishing, Cham, pp. 487–506.

Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004). Managing flexible products on a network, *Technical Report CORC TR-2004-01*, Computational Optimization Research Center, Columbia University.

Gilbert, F., Marcotte, P. and Savard, G. (2014a). Logit network pricing, *Computers & Operations Research* **41**: 291–298.

Gilbert, F., Marcotte, P. and Savard, G. (2014b). Mixed-logit network pricing, *Computational Optimization and Applications* **57**(1): 105–127.

Haase, K. (2009). Discrete location planning, *Technical Report ITLS-WP-09-07*, Institute of Transport and Logistics Studies, University of Sydney, Australia.

Haase, K. and Müller, S. (2013). Management of school locations allowing for free school choice, *Omega* **41**(5): 847–855.

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

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

Laporte, G., Nickel, S. and Saldanha da Gama, F. (2015). *Location Science*, Springer.

Liu, Q. and Van Ryzin, G. J. (2008). On the choice-based linear programming model for network revenue management, *Manufacturing & Service Operations Management* **10**(2): 288–310.

Marianov, V., Ríos, M. and Icaza, M. J. (2008). Facility location for market capture when users rank facilities by shorter travel and waiting times, *European Journal of Operational Research* **191**(1): 32–44.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142.

Müller, S., Haase, K. and Kless, S. (2009). A multiperiod school location planning approach with free school choice, *Environment and Planning A* **41**(12): 2929–2945.

Sharif Azadeh, S., Marcotte, P. and Savard, G. (2014). A taxonomy of demand uncensoring methods in revenue management, *Journal of Revenue and Pricing Management* **13**(6): 440–456.

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

Shen, Z.-J. M. and Su, X. (2007). Customer behavior modeling in revenue management and auctions: A review and new research opportunities, *Production and operations management* **16**(6): 713–728.

Snyder, L. V. (2006). Facility location under uncertainty: a review, *IIE Transactions* **38**(7): 547–564.

Talluri, K. T. and Van Ryzin, G. J. (2004). Revenue management under a general discrete choice model of consumer behavior, *Management Science* **50**(1): 15–33.

Talluri, K. T. and Van Ryzin, G. J. (2006). *The Theory and Practice of Revenue Management*, International Series in Operations Research & Management Science, Springer US.

van Ryzin, G. J. (2005). Future of revenue management: models of demand, *Journal of Revenue and Pricing Management* **4**(2): 204–210.

Vulcano, G. J., van Ryzin, G. J. and Chaar, W. (2010). OM practice - choice-based revenue management: An empirical study of estimation and optimization., *Manufacturing & Service Operations Management* **12**(3): 371–392.

Zhang, Y., Berman, O. and Verter, V. (2012). The impact of client choice on preventive healthcare facility network design, *OR spectrum* **34**(2): 349–370.

A Exhaustive results of the case study

A.1 Price calibration

R	Solution time	Prices PSP	Prices PUP	Demand PSP	Demand PUP	Demand FSP	Revenue
5	2.58 s	0.54	0.79	27.0	15.0	8.00	26.43
10	3.98 s	0.53	0.74	26.0	17.0	7.00	26.36
25	29.2 s	0.54	0.79	28.0	14.9	7.08	26.90
50	4.08 min (*)	0.54	0.75	25.2	17.8	7.00	26.97
100	20.7 min (*)	0.54	0.74	24.4	18.5	7.04	26.90
250	2.51 h (*)	0.54	0.74	24.8	18.2	7.03	26.85

(*) Instances not solved to optimality, gap of 0.01% for the MIP best bound found

Table 3: Solution time, optimal price, expected demand and expected revenue in the uncapacitated case

R	Solution time	Prices PSP	Prices PUP	Demand PSP	Demand PUP	Demand FSP	Revenue
5	12.0 s	0.63	0.84	18.2	17.2	14.6	25.91
10	54.5 s (*)	0.57	0.78	19.9	17.9	12.2	25.31
25	13.8 min	0.59	0.80	19.5	18.1	12.4	25.96
50	50.2 min (*)	0.59	0.80	19.5	18.2	12.3	26.10
100	6.60 h	0.59	0.79	19.1	18.7	12.2	26.03
250	1.74 days	0.60	0.80	19.0	18.1	12.8	25.93

(*) Instances not solved to optimality, gap of 0.01% for the MIP best bound found

Table 4: Solution time, optimal price, expected demand and expected revenue in the capacitated case

A.2 Price differentiation by population segmentation

Discount (%)	Solution time (min)	Prices R PSP	Prices R PUP	Demand R PSP	Demand R PUP	Demand R FSP	Prices NR PSP	Prices NR PUP	Demand NR PSP	Demand NR PUP	Demand NR FSP	Revenue
20	23.1	0.54	0.74	11.8	9.40	5.78	0.68	0.92	7.46	8.60	6.94	29.7
25	54.2	0.51	0.69	12.2	10.2	4.64	0.68	0.92	7.34	8.72	6.94	30.7
30	51.3	0.50	0.67	12.7	10.4	3.86	0.72	0.96	6.16	8.50	8.34	31.8
40	63.9	0.48	0.65	13.7	10.7	2.60	0.80	1.08	4.88	7.20	10.9	34.2
50	67.3	0.46	0.64	15.0	10.4	1.62	0.92	1.28	3.74	5.32	13.9	37.3

Table 5: Solution time, optimal price and expected demand by segment and expected revenue for different subsidies for approach 1

Discount (%)	Solution time (min)	**Prices R**		**Demand R**			**Prices NR**		**Demand NR**			
		PSP	PUP	PSP	PUP	FSP	PSP	PUP	PSP	PUP	FSP	Revenue
20	44.1	0.54	0.74	11.8	9.40	5.78	0.68	0.92	7.46	8.60	6.94	26.3
25	24.6	0.51	0.69	12.2	10.2	4.64	0.68	0.92	7.34	8.72	6.94	26.2
30	35.7	0.50	0.67	12.7	10.4	3.86	0.72	0.96	6.16	8.50	8.34	26.0
40	130	0.46	0.62	13.5	11.2	2.30	0.76	1.04	5.88	7.36	9.76	25.3
50	174	0.42	0.58	14.1	11.8	1.08	0.84	1.16	4.76	6.18	12.1	23.9

Table 6: Solution time, optimal price and expected demand by segment and expected revenue for different subsidies for approach 2