Transportation Research Part E 195 (2025) 103993

Contents lists available at ScienceDirect

Transportation Research Part E

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

Enhancing carsharing pricing and operations through integrated choice models

Beatriz Brito Oliveira [a,*], Selin Damla Ahipasaoglu [b]

[a] INESC TEC, Faculty of Engineering, University of Porto, Portugal
[b] CORMSIS, School of Mathematical Sciences, University of Southampton, United Kingdom

ARTICLE INFO

Keywords:
Carsharing
Pricing
Fleet relocation
Combinatorial optimization
Discrete choice models

ABSTRACT

Balancing supply and demand in free-floating one-way carsharing systems is a critical operational challenge. This paper presents a novel approach that integrates a binary logit model into a mixed integer linear programming framework to optimize short-term pricing and fleet relocation. Demand modeling, based on a binary logit model, aggregates different trips under a unified utility model and improves estimation by incorporating information from similar trips. To speed up the estimation process, a categorizing approach is used, where variables such as location and time are classified into a few categories based on shared attributes. This is particularly beneficial for trips with limited observations as information gained from similar trips can be used for these trips effectively. The modeling framework adopts a dynamic structure where the binary logit model estimates demand using accumulated observations from past iterations at each decision point. This continuous learning environment allows for dynamic improvement in estimation and decision-making. At the core of the framework is a mathematical program that prescribes optimal levels of promotion and relocation. The framework then includes simulated market responses to the decisions, allowing for real-time adjustments to effectively balance supply and demand. Computational experiments demonstrate the effectiveness of the proposed approach and highlight its potential for real-world applications. The continuous learning environment, combining demand modeling and operational decisions, opens avenues for future research in transportation systems.

1. Introduction

Balancing the supply and demand dynamics in carsharing systems is a complex task, offering both challenges and opportunities for effective and innovative management. Carsharing systems are characterized by a centrally owned and managed fleet of vehicles, which is shared among different users that can rent the cars for a short period. These systems are highly dependent on operational efficiency to provide a reasonable service level to their users and thus ensure the system's economic viability. There are different configurations of carsharing systems, with two key differences being the existence (or not) of previously defined parking areas or stations and the possibility (or not) of returning the car to the same location from which it was picked up. These configurations are called station-based systems if the users must pick up and drop off the car at a given station. Within these, we refer to round-trip systems - where the drop-off point must be the same as the pick-up point - or one-way systems - where the user can choose the drop-off point. Systems not bounded by previously defined stations are called free-floating one-way systems, and users may pick up and drop off the car at any point within the system's operating area, which makes them the most flexible and user-friendly type of carsharing system (Shaheen et al., 2015). This, unfortunately, also makes them more challenging to operate efficiently.

* Correspondence to: Faculty of Engineering, University of Porto – Rua Dr. Roberto Frias s/n, 4200-465, Porto, Portugal
E-mail address: beatriz.oliveira@fe.up.pt (B.B. Oliveira).

https://doi.org/10.1016/j.tre.2025.103993
Received 3 July 2024; Received in revised form 20 December 2024; Accepted 21 January 2025
Available online 30 January 2025
1366-5545/

Especially in this setting, operational decisions on short-term price promotions and fleet relocations are critical to match supply and demand. Fleet relocations – also known as operator-based relocations – refer to the process of strategically moving vehicles to ensure better availability and accessibility for users. Fleet relocations are essential to demand management in carsharing systems as they involve the strategic adjustment of vehicle distribution to align with varying demand patterns, often from low demand areas to high demand areas. Carsharing operators use this proactive approach to fleet management to help meet user expectations, reduce unmet demand, and ultimately contribute to the overall efficiency and effectiveness of the carsharing service. Nevertheless, relocations come at an economic and environmental cost that should be considered.

Relocation operations in carsharing are notably different from those in other shared mobility modes. For example, in bikesharing (and similar micro-mobility systems), relocations are decided by the operator and often performed in batches utilizing larger vehicles. Conversely, in ride-hailing platforms (and other driver-based shared mobility systems), while recent research is also focused on dynamic pricing and relocation strategies (Chen et al., 2024), drivers are strategic agents who can relocate their vehicles in anticipation of future demand and prices. In carsharing, relocations are decided for each car individually and they require a staff member to move the car. However, the operator can also employ other incentive mechanisms, typically price reductions, to give incentives to the users to drop the car off in convenient locations in order to reduce relocations by hired drivers. These incentives are sometimes referred to as user-based relocation. The following example illustrates how the two relocation methods available in carsharing systems are intrinsically connected and also how they relate to the importance of accurately predicting and modeling demand.

Let us assume that, at a certain period of time, the carsharing operator will need one additional car to fulfill the expected demand in a given area. The expected revenue from that extra trip is 4 monetary units. The three options for the operator are: (1) to miss that potential rental and the associated revenue, (2) to have a staff member relocate the car from an area with an oversupply of vehicles, with an estimated cost (including staff time and fuel cost, as examples) of 3 monetary units, or (3) to offer a discount to a user that would otherwise drop off its car in a near-by oversupply area. Since the costs to relocate a vehicle are smaller than the expected revenue of this trip, it is beneficial to balance the system this way. The best relocation method – operator-based (option 2) or user-based (option 3) – will depend on the discount offered. In this example, if a discount of up to 3 monetary units were enough to incentivize a current user's change in behavior, which is inherently uncertain, it would be best to select option 3.

While operators of different shared mobility systems have common objectives, face similar challenges, and use pricing incentives to manipulate the demand and the distribution of assets, they differ significantly at the operational level in terms of the number of agents, control variables, and decisions to be made. We study an operational problem that considers car movements (either by users or by operator-based relocations) within a time-space network, where the space dimension covers the operating area and time is discretized over a given time horizon. Each trip type has specific pick-up and drop-off times, locations, and an associated expected demand. This is a relevant and challenging problem where multiple products (trips by users) share the same resources (homogeneous fleet), and the time and space positioning of the vehicles impacts the efficiency of the system and the service provided.

Understanding the underlying choice behavior of the users is critical for operating these systems. As carsharing is competing with other modes of transportation, even though the price is not the only attribute that impacts the user's decision, it plays a significant role (Huang et al., 2018). Discrete Choice Models model the observed and latent relation of price and other relevant attributes to demand (Ben-Akiva and Bierlaire, 1999), and therefore allow the development of a prescriptive tool that can be used to make operational decisions to "shape" demand to match the most efficient car relocation patterns. In a nutshell, we demonstrate that a choice model that is trained by aggregated information from different types of trips, can be embedded in these decision-support models to improve operational decisions and overall results.

In this problem, we assume that the free-floating one-way carsharing system operates within a city that is divided into smaller zones. Demand arrives at the time of the trip with no previous reservation. We assume that we know the total potential demand for a given trip type (i.e., a trip starting in a given pick-up zone and time and ending in a given drop-off zone). This assumption is based on the fact that typical and aggregated daily movements of people between city zones without modal split can be estimated from different sources and population surveys (Jorge et al., 2015). We estimate the expected share of the potential demand that uses the carsharing system using a binary logit model based on the trip attributes. The trip price is charged per minute, depending on the pick-up area and time. Baseline prices are set and known to the users a priori. Nevertheless, the operator can use promotions in specific areas and times of the day as a mechanism to encourage a more favorable distribution of cars over the network. Therefore, users can benefit from time-limited promotions, depending on the pick-up area and time. Additionally, the operator can perform empty transfers or relocations (i.e., the car is driven by a staff member to another location) at any time, at a given cost. These assumptions are realistic with the current practice as discussed in Golalikhani et al. (2021). Our modeling framework aims to maximize the operator's expected contribution margin (i.e., considering variable revenues and costs) during a given time horizon (e.g., one week) by identifying the optimal combination of promotions and relocations. This often also helps reducing associated carbon emissions.

The main contributions of this work are twofold: the first is related to the operational challenges of balancing supply and demand in free-floating one-way carsharing systems, which is enabled by the second contribution, which is related to the reduced data requirements on the estimation of the choice model by aggregation of information collected from all trip types. **First**, we propose a mixed-integer linear programming (MILP) model for this pricing and operations problem that embeds a binary logit model to estimate demand for carsharing trips. Even though binary logit models are often used for demand modeling in transportation systems, they are typically applied in a two-stage process, where estimation precedes (and is independent of) prescription. This is perhaps because they bring additional complexity to the model and increase the time to solve. We propose to use the binary logit model not only for demand estimation but also as a mechanism to shape and manage the demand in an integrated system. The binary logit-based mathematical programming model proposed obtains reduced optimality gaps within reasonable times, making it applicable in reality. **Second**, in the few cases where discrete choice models have been embedded in operational decision-making, they are often only based on user costs (or price). We propose a utility model where the pick-up time, pick-up area and drop-off area are also features of the trip, which allows the integration of critical information related to spatial and temporal properties. This information is acquired from different trip types over the space–time network over time. To avoid a high number of features, we propose a categorizing strategy for different trip areas and periods. In this context, we demonstrate (i) how integrating multiple trips in a single binary choice demand model (in which properties of the trip type are features) improves estimation and (ii) how categorizing strategies allow for reducing computational times while keeping a similar estimation performance. The proposed framework is further validated by the analysis of a carsharing trip dataset from Milan, Italy. Overall, the proposed modeling framework opens new avenues of research in improving operational decisions in this context through better demand management.

The remainder of the paper is structured as follows. Section 2 discusses the literature on carsharing pricing and operations, and demand estimation with discrete choice models, clearly positioning the contributions of this work. The modeling framework and the corresponding mathematical models are presented in Section 3. Section 4 details and discusses the computational experiments that show the promising potential of the proposed approach, and, finally, Section 5 summarizes the contributions and limitations of this work and proposes future research directions opened by this research such as extending the demand modeling to account for demographics of users and consideration of multi-mode travel systems in which carsharing is one of the modes considered. Beyond addressing operational challenges in one-way carsharing, the second contribution of this work, which emphasizes reduced data requirements and utility modeling for demand estimation, has implications for improving decision-making in other transportation applications like bus ticketing and congestion pricing.

2. Literature review

Research in carsharing pricing and relocation has become active in the past years, due to its connections to societal and environmental challenges such as reducing carbon emissions, boosting the shared economy, and ensuring economic viability and environmental sustainability with limited resources. Integrating these two mechanisms – pricing and relocation – is increasingly recognized as critical to balance supply and demand mismatches in one-way carsharing systems. In addition, demand modeling and in particular understanding the relationship between price and demand in the presence of other factors such as the locality and time of the trips play a critical role in these models. In this section, we will focus on recent advances in this field, highlighting their contributions and identifying the gaps we aim to bridge with our work.

Relocation and pricing are critical for carsharing operations. Designing and operating a carsharing system encompasses several strategic, tactical, and operational decisions from the operator that often interconnect with the decisions of potential users. The two main approaches in the literature to tackle these problems are optimization and simulation. Optimization approaches focus on finding the best solution within feasibility boundaries, while simulation approaches model real-world processes by generating and analyzing multiple scenarios to assess their performance. We refer to Golalikhani et al. (2021) for a detailed review of the topic of decision-making for carsharing operators.

Jorge et al. (2014) demonstrate that relocations are critical for the profitability of one-way station-based carsharing systems, even when pricing is not used for demand management. The authors propose an optimization model and a metaheuristic for its solution, and validate their impact using real data from Lisbon, Portugal, including potential station locations and average number of trips. Boyacı et al. (2017) propose integrated optimization-simulation framework for vehicle and staff relocations for one-way electric carsharing systems (assuming there might be reservations in advance). The approach is validated with a large scale real-world dataset and different policies for serving requests are investigated. Recently, Hosseini et al. (2024) propose a relocation policy for carsharing based on a fluid model approximation of the dynamic problem. For a full review of the vehicle relocation problem in one-way carsharing systems, as well as its assumptions, models and solution methods, we refer to Illgen and Höck (2019). In Jorge et al. (2015), trip pricing is analyzed as a strategy (independent from relocation) to balance the fleet across the system. Also using data from Lisbon, the authors demonstrate that optimizing prices is key for profitability, yet this leads to higher prices for users and lower service levels. Integrating pricing with relocation has potential to overcome this issue.

Even though the literature on decision-support methods for carsharing operations initially focused on relocation and pricing separately, there are several recent attempts for their integration with promising results including the work of Lu et al. (2021). The authors propose a bilevel model that maximizes the operator's profit at the upper level by deciding on pricing and relocations, while, at the lower level, users select the travel mode, minimizing their cost. Huang et al. (2021) also tackle the carsharing pricing and relocation simultaneously, considering the additional feature of *access trips*: when the user walks or rides a bike to pick up the shared car. The authors propose a two-stage stochastic model where pricing is the first-stage decision (made before demand uncertainty is realized), and relocations are the recourse decisions to react to the different realizations of demand uncertainty. The model is validated using data from Suzhou, China, and the results demonstrate the profitability of the integrated approach and the added value of flexibility due to the availability of access trips in the model. Pantuso (2022) proposes an exact solution method for the two-stage stochastic carsharing pricing and relocation problem and validates the approach using data from Milan, Italy, with randomly generated demand. The pricing decision consists of optimizing the drop-off fee for each area. This drop-off fee is added to a fixed per-minute fee, which is independent of origin or destination and is a given parameter of the model, unlike to the model we propose. In this context, the concept of *user-based relocation* arises, where the operator offers incentives to users to relocate cars. This mechanism is parallel to using pricing as an incentive. Huang et al. (2020) compare the efficiency of user- and operator-based relocation methods in a one-way station-based electric carsharing system, showing that the imbalance problem can be well addressed by both operations.

Banerjee et al. (2021) provide a general approximation framework to optimize shared vehicle systems by addressing spatial supply–demand externalities using steady-state Markovian models, showing that queueing theory has also been leveraged as a robust tool for modeling shared mobility systems. In Benjaafar et al. (2023), the importance of a learning component in this setting is discussed, and the authors propose an online learning algorithm for pricing in on-demand vehicle-sharing networks, ensuring convergence to optimal static pricing policies. Braverman et al. (2019) explore the issue of (driver-based) relocation in ride-sharing systems, showing that fluid-based optimizations can effectively manage fleet availability and improve system performance in these large-scale markets. These studies highlight how demand learning is key to dynamically adjusting pricing and managing supply to balance the network. By integrating demand learning with car relocations, novel insights can be achieved to improve supply–demand interactions, ensuring higher vehicle availability and enhanced operational efficiency. To achieve this in the setting analyzed in this paper, we focus on mathematical models and solution techniques that are fit for analyzing and operating smaller-scale carsharing systems, where user-based relocation works alongside traditional operator-based relocations to improve the system's efficiency.

The quality and effectiveness of the proposed operational models are significantly affected by the assumptions used to model the demand-price relationship. Nevertheless, there is still no consensus in the literature on the best approach in this context. Ren et al. (2019) develop a trip pricing model for one-way carsharing systems with Electric Vehicles (EV). EV carsharing poses additional challenges, requiring integration with power systems (vehicle-grid integration). The authors model demand according to a linear function where a price elasticity for a given price is multiplied by a base demand level. Wang and Ma (2019) propose a pricing model to influence demand, where the price-demand relationship considers price elasticity and accounts for potential changes in the origin and destination stations influenced by pick-up and drop-off specific rewards. In the work by Pantuso (2022) described before, to model demand, a utility function for carsharing is used (linear with the pricing decision variable of the model) where the unknown part of the utility is stochastic and discretized. Additionally, several other competing modes have a given utility (exogenous to the model). The demand model assumes each user chooses the alternative with the highest utility. Overall, linear price-demand models with different levels of detail are frequent in the literature. Linear price-demand models are not designed to capture complex behavioral patterns such as product substitution and nonlinear relationships between demand and product attributes, and therefore their applicability is limited in practical applications. On the other hand, discrete choice models (DCMs) (McFadden, 1981) account for substitution while also being able to use various attributes of the user and the service offered, including but not limited to its price. These characteristics often make DCMs the preferred method for transportation researchers in other applications, such as route planning and congestion pricing. See Ben-Akiva and Bierlaire (1999) and Ortúzar and Willumsen (2011) for an extensive and insightful discussion on the use of DCMs for demand estimation in transport.

Huang et al. (2018) tackle an integrated problem for one-way station-based carsharing systems where the proposed model decides on the strategic location of stations while considering relocations. The authors use the logit DCM for modeling carsharing utility versus private car utility for each origin and destination. In this model, carsharing utility considers as features the trip's price (a parameter in this setting), walking and travel time costs, whereas private car utility considers fixed vehicle costs, parking and travel time costs. In the linear-in-parameters utility functions, the weights of all cost components (four for each utility function) are identical. The authors validate their approach using data from Suzhou, China, with randomly generated travel demand. A similar approach is applied in the bilevel model described above from Lu et al. (2021). Demand also depends on a logit model with two alternatives — carsharing and private car. In this case, carsharing utility depends only on an alternative-specific coefficient and the trip's price. In our work, we build on the previous papers that integrate carsharing pricing and relocation, and propose a new demand management model to be embedded in the operational/prescriptive framework. In particular, as will be explained in Section 3.1 we estimate demand with a logit DCM that relies on trip-related features besides cost, such as the trip's starting time, allowing for specific trips with little information (due to limited number of observations of trips with identical features) to be estimated more accurately based on data collected from similar trips. Liu et al. (2022) study an interesting approach to user-based relocation, where the users' willingness to relocate (i.e., to accept an incentive to be flexible with their drop-off location) is learned from a data-driven, dynamic approach, based on a binary logit model. The authors propose some extensions to this approach and validate it with numerical experiments based on data from Singapore. There are several common elements between our approach and the methodology of Liu et al. (2022) such as the dynamic learning environment, using relocations and promotions as operational tools, and building a logit based utility function to capture the stochastic demand. Nevertheless, the two papers differ significantly in the modeling and solution of each of these elements. In particular, Liu et al. (2022) focus on real-time decisions considering uncertainties in trip requests and travel time, use an approximate dynamic programming formulation for decision-making, and a Bayesian framework to update model parameters. In the utility specification, only the characteristics of the users are considered regardless of the time of the day, and the pick-up and drop-off locations of the trip that is being considered. In our case, we use an optimization model to make key decisions such as relocations and prices of each trip over the space–time network in which the demand model is integrated. In the utility specification of the demand model, we take trip characteristics into account and integrate information gathered from various trips over different time intervals and locations. Our demand estimation model can easily be extended to include customer characteristics, e.g., using a mixed logit model and customer segmentation. It might also be possible to add trip characteristics to the model of Liu et al. (2022).

Finally, other relevant and interesting extensions are currently being explored in the carsharing pricing and relocation literature. The first extension relates to carsharing competition. In several carsharing systems, only one operator offers this mobility service in the same city or region (or one operator with a substantial market share can act as a monopolist). In other cases, there is competition between carsharing operators. This setting has a significant impact on pricing strategies. Balac et al. (2019) study this

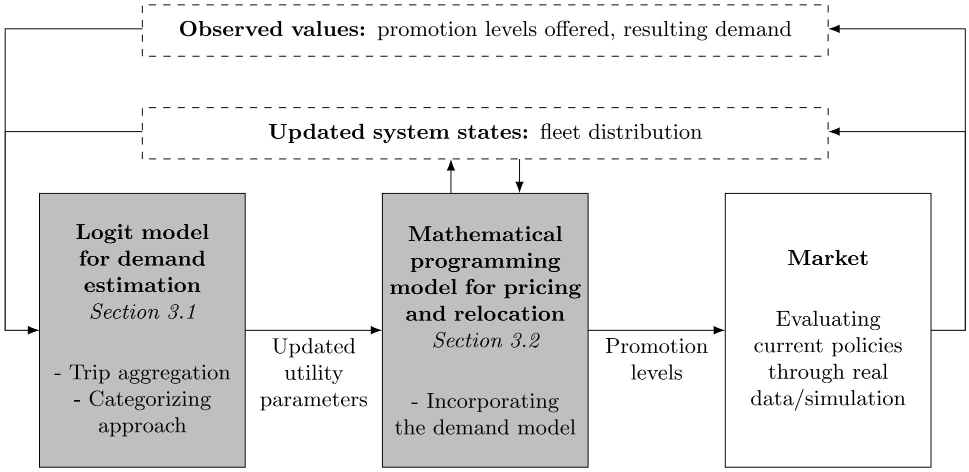


Fig. 1. Modeling framework.

problem, analyzing different pricing (and relocation) scenarios for Zurich, Switzerland. Yang et al. (2022) also tackle the pricing and relocation problem in a competitive market and propose a multi-leader-following game model, validated with data from Quanzhou, China. Another relevant challenge arises within the carsharing pricing and relocation problem when freelance drivers are hired to perform the relocations instead of the staff from the carsharing operator (Samie and Rezaee, 2022). This is translated directly into traditional models by converting the relocation cost into a decision variable (i.e., the freelance driver compensation). The authors propose a mixed integer model for this setting, showing that combining these strategies benefits the operator. They assume a linear price-demand relationship. Finally, most literature on one-way carsharing pricing and relocation, even if tackling a free-floating system, depends on key assumptions adapted from station-based models, namely regarding the discretization of the operating area and of the considered time period. This could be especially relevant for matching supply and demand, considering, e.g., the walking distance. Soppert et al. (2023) investigate this issue and demonstrate the advantages of accurately accounting for free-floating flows in the operator's profitability. Our work can be adapted for such systems easily by careful selection of relevant features.

As discussed throughout this chapter, the approach proposed in this paper differs from the above-mentioned papers regarding objectives, scope, and methods, and it complements the previous research. In summary, this work presents an approach where pricing and relocation are optimized simultaneously in an integrated model. This differs from two-stage approaches, where pricing is decided before demand uncertainty is revealed, and relocations are the recourse decisions that allow reacting to different demand levels. The proposed solution approach is thus suited to be applied in practice, iteratively and on-the-fly, using accumulated data. Additionally, we optimize prices that depend on origin, destination, and starting time, following the complexity observed in practice. At the same time, we model the uncertain relationship between demand and price using a discrete choice model that is embedded in the operational framework. This model relies on trip-related features besides cost, allowing for specific trips with a limited number of observations to be estimated more accurately based on data collected from similar trips.

3. Carsharing pricing and relocation models

This section presents the mathematical models proposed for carsharing pricing and relocation, including the demand estimation and management components. To tackle the problem presented in this paper, we propose a modeling framework that involves several relevant components that will be discussed thoroughly in the following sections. Fig. 1 presents this modeling framework, highlighting the relationship between its components.

Throughout this paper, demand modeling is based on a binary logit model described in Section 3.1. This model is characterized by aggregating different trips under one utility model (with two alternatives: using or not using the carsharing alternative), thus allowing for information on similar trips to improve estimation. Additionally, we propose a categorizing approach to reduce the number of parameters to be estimated. This has two benefits, the number of required data points and the computational time to fit the demand model are both reduced without compromising the model accuracy. The modeling framework relies on a dynamic structure where the binary logit model is applied to estimate demand using the accumulated observations of past iterations at the start of each decision point. Aggregating categorical variables such as location and time into a small number of categories reduces the number of parameters to be estimated and therefore helps getting accurate models even when the number of observations is limited such as early in the time horizon. At the core of the modeling framework is a mathematical program that prescribes the optimal promotion levels to offer users and relocations to match supply and demand (Section 3.2). This modeling framework is based on a continuous learning environment where the responses of the market to the decisions taken are observed and accumulated to improve estimation and decision-making dynamically.

3.1. Logit model for demand estimation

We assume that users of the carsharing system are homogeneous and arrive at the carsharing platform with the intention of booking one car at a time for a trip type (specified with its destination and starting time) and observe the associated price. At this point, they effectively choose between two alternatives: using a shared car with the offered price or leaving the platform without using the shared car (i.e., choosing the outside option). Price is one of several factors that may influence users' mode choice decisions, along with other significant factors such as convenience, travel time, and service attributes. Here, we focus on price as the key demand-influencing factor and use price reductions as incentives. This is in line with the common practice. We also include attributes corresponding to the time of the day and the location of the demand. Travel time is not included since it is assumed to be fixed for each pair of locations and captured implicitly in our model. This means that we work with a relatively simple model, which captures the essential components for our study, but remains to be easy to estimate and interpret. We assume the users need to take the trip at a specific time and therefore do not change the time of their trip by exploring time-dependent prices, which is also a common assumption in dynamic pricing literature.

The utility of using a shared car is modeled as in Eq. (1), where z is the deterministic utility and $\tilde{\varepsilon}$ represents the error term which follows the standard Gumbel distribution. The deterministic utility z is defined by the sum of an alternative-specific coefficient and the product of a vector of coefficients by a vector of observable attributes of the trip (see Section 3.1.)

$$u = z + \tilde{\varepsilon} \tag{1}$$

The utility of the outside alternative is set to zero; therefore, the carsharing utility assumes a relative value compared with competing modes of transportation. We assume that users choose the alternative that maximizes their utility, and therefore obtain the standard binary logit model in which the choice probability of the carsharing option is given by $\frac{\exp(z)}{1+\exp(z)}$.

We consider that the city is divided into a set of areas $\mathcal{A}$, and the time horizon is discretized in a set of time periods $\mathcal{T}$. We estimate demand for each *trip type*, which is defined as a specific combination of the trip attributes: the pick-up area $i \in \mathcal{A}$, drop-off area $j \in \mathcal{A}$, and pick-up period $t \in \mathcal{T}$ using the logit model described above. We assume that the duration of the trips is fixed according to the pick-up and drop-off areas, as in Huang et al. (2018).

Assuming no knowledge of the user's socio-demographic characteristics, we propose using the characteristics of the trip (pick-up area and period, and drop-off area) and its final price (i.e., after the promotion has been applied) as the relevant attributes that determine the utility that is given to the carsharing alternative. Note that the input data for these attributes are nominal. Therefore, we need to employ an appropriate "one-hot-encoding" step when building the linear-in-parameter model of z, resulting in a large number of model parameters. In order to reduce the number of parameters to be estimated, we assume that the values of these features can be grouped into categories according to their similar characteristics (e.g., peak vs off-peak periods or high-density vs low-density areas). The goal is to group time periods that have a similar effect on users' propensity to use carsharing and areas that have a similar attractiveness as pick-up or drop-off points. The areas and time periods can be categorized using any common partitioning technique, as will be shown later in this paper. We introduce the following notation to capture this:

- category of area i: $G(i) = \{1, \dots, M\}$ — discrete, non-ordered set of M area categories,
- category of pick-up time t: $B(t) = \{1, \dots, N\}$ — discrete, non-ordered set of N time categories,
- trip final price: price_{ijt} — price per period and promotion depends on pick-up area i and period t; however, the duration of the trip depends on the drop-off area j, thus influencing the trip's final price (the calculation of the final price is detailed in Section 3.2).

Note that M, the number of area categories, should be less than or equal to the cardinality of the set of areas $\mathcal{A}$ (and the same relationship holds for N and $\mathcal{T}$) since the categories aggregate the larger number of specific areas and periods. For representing the discrete, non-ordered sets of categories, we apply the popular one-hot encoding technique, i.e., auxiliary binary variables are defined for time categories ($m_{B(t)}$), pick-up area categories ($n_{G(i)}$) and drop-off area categories ($l_{G(j)}$). Table 1 illustrates the use of these variables. To better understand this, let us focus on the attribute of time. Time is naturally a continuous variable but it is often discretized for practical reasons. Even though discrete time periods follow a (time) order, since their impact on demand is not dependent on that order, they may be seen as nominal (rather than ordinal) data. Trip 1 starts at 8 AM and trip 4 starts at 6 PM. These pick-up times are separated (one is in the morning and the other in the afternoon). Nevertheless, they are both peak periods where demand is the highest in the system. Therefore, in our approach, they fall within the same pick-up time category ($B(8AM) = B(6PM) =$ Peak). The auxiliary variables needed for the one-hot encoding of these categories are as follows and their application to the previous example is found in columns 9–14 in Table 1:

$m_{B(t)} = \{0, 1\}, \forall B(t) = \{1, \dots, N-1\}$ — $= 1$ if the carsharing trip has its pick-up at a time that belongs to category $B(t)$, $= 0$ otherwise, with $m_N = 1 - \sum_{B(t)=1}^{N-1} m_{B(t)}$.

$n_{G(i)} = \{0, 1\}, \forall G(i) = \{1, \dots, M-1\}$ — $= 1$ if the carsharing trip has its pick-up at an area that belongs to category $G(i)$, $= 0$ otherwise, with $n_M = 1 - \sum_{G(i)=1}^{M-1} n_{G(i)}$.

$l_{G(j)} = \{0, 1\}, \forall G(j) = \{1, \dots, M-1\}$ — $= 1$ if the carsharing trip has its drop-off at an area that belongs to category $G(j)$, $= 0$ otherwise, with $l_M = 1 - \sum_{G(j)=1}^{M-1} l_{G(j)}$.

In summary, Eq. (2) describes the deterministic part of the carsharing utility function.

$$z = \alpha + \sum_{B(t)=1}^{N-1} \left(\beta^1_{B(t)} \cdot m_{B(t)}\right) + \sum_{G(i)=1}^{M-1} \left(\beta^2_{G(i)} \cdot n_{G(i)}\right) + \sum_{G(j)=1}^{M-1} \left(\beta^3_{G(j)} \cdot l_{G(j)}\right) + \beta^4 \cdot \text{price}_{ijt}, \tag{2}$$

Table 1
Example of the deterministic part of the utility function using auxiliary binary variables and categories to aggregate pick-up and drop-off areas and pick-up periods into two categories each: “Peak” and “Low”.

							Attribute variables							
								$m_{B(t)}$		$n_{G(i)}$		$l_{G(j)}$		Price
Trip	i	j	t	$B(t)$	$G(i)$	$G(j)$		m_{Peak}	m_{Low}	n_{Peak}	n_{Low}	l_{Peak}	l_{Low}	
1	2	7	8 AM	Peak	Peak	Peak		1	0	1	0	1	0	10
2	5	8	8 AM	Peak	Peak	Low		1	0	1	0	0	1	15
3	5	9	3 PM	Low	Low	Peak		0	1	0	1	1	0	10
4	9	1	6 PM	Peak	Low	Low		1	0	0	1	0	1	5
...	...					...	...							...
					Related Parameters		α	β^1_{Peak}	β^1_{Low}	β^2_{Peak}	β^2_{Low}	β^3_{Peak}	β^3_{Low}	β^4

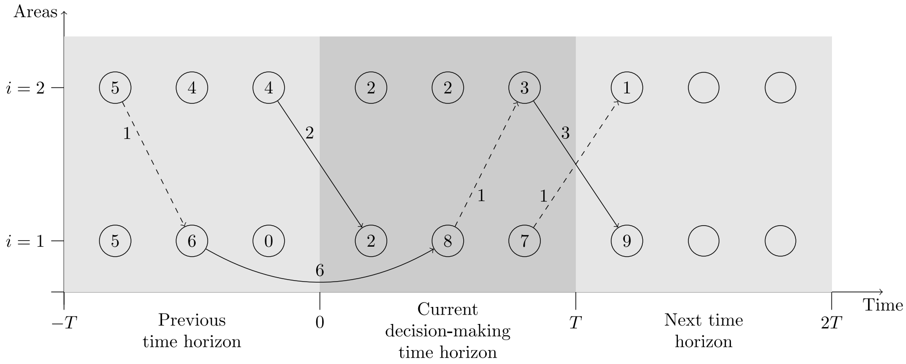


Fig. 2. Example of how decisions from different time horizons interact, with two areas and three time periods within each decision-making time horizon, and considering a fleet of 10 cars. The system is a time-space network where the number associated with each area/period is the number of cars available. Solid lines represent trips by users and dashed lines represent relocations, with the associated number representing the number of cars in each.

where we have dropped one of the binary attributes in each hot-encoded categorical variables to avoid dependency (multicollinearity) between features used in the model as common in practice.

3.2. Mathematical programming model for pricing and relocation

We propose a modeling framework with three main components: the operational setting, mostly related to fleet management issues; the pricing setting, concerned with the pricing policies and assumptions; and the choice modeling (or demand estimation) setting, where demand-related assumptions and modeling choices are detailed and explained. We consider a decision-making time horizon for which the operational decisions (prices and relocations) are under considerations. The parameters for the current decision-horizon, such as the distribution of available cars at each time slot, naturally depend on decisions made in previous time horizons. This is exemplified in Fig. 2 for a toy example and discussed below in more detail. The notation used and the modeling framework, especially for the operational setting, are mostly based on Oliveira et al. (2018).

Operational setting:. As explained before, we consider that a city is divided into a set of areas $\mathcal{A}$, and the time horizon is discretized in a set of periods $\mathcal{T}$. Throughout this work, we refer to trip type as a trip taking place between a pick-up area $i \in \mathcal{A}$ and drop-off area $j \in \mathcal{A}$, starting at pick-up period $t \in \mathcal{T}$. We assume the duration of the trips depends on the time to drive (denoted as K_{ijt}) between the pick-up area i and the drop-off area j, which also depends on the pick-up time t to account, e.g., for changing traffic conditions. We assume relocations take the same time as carsharing trips with the same pick-up and drop-off areas, and same starting time. Additionally, we consider C_{ijt} to be the cost of relocating a car between areas i and j starting at time t. When a time horizon for optimization starts, the system is already running. Therefore, some cars are idle and waiting to be rented, and their current deployment results from trips in the previous time horizon. In the example in Fig. 2, area $i = 2$ starts the current decision-making horizon with 2 cars available since trips and relocations in the previous time horizon reduced the number of cars available. Similarly, some ongoing trips and relocations (decided in the previous time horizons) will increase the capacity at a certain point in the current time horizon (and at a particular location). In Fig. 2, for example, six trips from users starting at area $i = 1$ in the previous time horizon are scheduled to arrive to the same area during the current time horizon, increasing the number of cars available. Previous decisions thus determine the system’s starting state, impacting the optimization of current decisions, which is limited by these conditions. To acknowledge this, we assume that the system is running at the beginning of the decision horizon with a given initial number of idle cars (W_i) at each area i. Throughout the horizon, we acknowledge that some cars that are currently being relocated (R_{it}), as well as some cars currently on ongoing user trips (O_{it}) will arrive at area i at time t. These parameters result from decisions made in the previous time horizons and impact the decisions made in this horizon. This leads to the need to test these models iteratively in a rolling horizon to avoid effects related to the start and end of the horizon.

Pricing setting:. As mentioned, the final price for carsharing trips depends on a base price per period, the promotion level we decide to offer, and the trip duration. For increased realism, we assume the base price is calculated step-wise. For example, for the first 10 min, the cost per minute might be higher than the cost for the following minutes of the trip. To model this, we assume the carsharing operator has established a set of steps $\mathcal{S}$ and knows:

- The length (in periods) of each pricing step $s \in \mathcal{S}$ (L_s).
- The last pricing step for a trip starting in area i at period t and ending at j (Q_{ijt}). This step can be directly derived from the total time of the trip (K_{ijt}).
- The base price (in monetary units) charged per period for trips starting in period t and area i in pricing step s (B_{its}).
- The promotion value (percentage of the full price) associated with promotion level $p \in \mathcal{P}$ for period t and area i (V_{it}^p), where $V_{it}^0 = 0$.

Therefore, the full price (F_{ijt}) depends on the step-wise base price per period and the duration of the trip, such that:

$$F_{ijt} = \sum_{s=1}^{Q_{ijt}-1} B_{its} \cdot L_s + B_{i,t,Q_{ijt}} \cdot (K_{ijt} - \sum_{s=1}^{Q_{ijt}-1} L_s). \tag{3}$$

The full price is a parameter of the mathematical programming model since the decision is on the promotion (or discount) offered (if any). The final price charged to the user, denoted as price_{ijt} in Eq. (2), depends on F_{ijt} and the promotion decision variables, as will be detailed in the following paragraph (see Eq. (6)).

Demand modeling:. In this setting, we assume that we know the total potential demand for a trip starting in period t and area i and ending in area j (denoted as D_{ijt}). Based on the utility function described in Section 3.1 (see Eq. (2)), applying a binary logit choice model, for a trip starting in area i and time t and ending in area j, we estimate the share of potential demand that selects the carsharing alternative with the following expression:

$$\text{share} = \frac{\exp(z_{ijt})}{\exp(z_{ijt}) + \exp(0)}. \tag{4}$$

To include this expression in the model linearly, we introduce parameter Z_{ijt}^p, which may be calculated for each promotion level and trip before running the model. It corresponds to the probability of a representative user choosing carsharing for each promotion level $p \in \mathcal{P}$ offered for trips starting at period t in area i going to area j. The parameter is calculated based on the categorizing approach for areas and periods described in Section 3.1. It is detailed in Eq. (5) and (6), where z_{ijt}^p represents the value of the deterministic part of the utility of a given type of trip for a given promotion level offered.

$$Z_{ijt}^p = \frac{\exp(z_{ijt}^p)}{\exp(z_{ijt}^p) + \exp(0)} \tag{5}$$

$$z_{ijt}^p = \alpha + \sum_{B(t)=1}^{N} \left(\beta_{B(t)}^1 \cdot m_{B(t)}\right) + \sum_{G(i)=1}^{M} \left(\beta_{G(i)}^2 \cdot n_{G(i)}\right) + \sum_{G(j)=1}^{M} \left(\beta_{G(j)}^3 \cdot l_{G(j)}\right) + \beta^4 \cdot F_{ijt}\left(1 - V_{it}^p\right) \tag{6}$$

The notation used for indices and parameters is summarized in Appendix A.

Decision variables

q_{it}^p	$\begin{cases} 1, \text{ if promotion level } p \text{ is charged for trips starting at time period } t \text{ in area } i \\ 0, \text{ otherwise} \end{cases}$
y_{ijt}	Number of cars relocated at time t from area i to area $j \neq i$
Auxiliary decision variables:	
x_{it}	Number of cars in area i at time t
w_{ijt}^p	Number of fulfilled user trips starting in time period t and area i and ending in area j, with promotion level p
r_{ijt}	Probability of a user selecting the carsharing alternative for a trip from area i to j in period t

The objective function of the model will account for the expected revenue, which is the result of multiplying the expected number of user trips and their final price. Since both of these elements depend on decision variables, to avoid a non-linear objective function, we add an index for the promotion level (p) in the decision variable that captures the share of the demand for the given trip type (i.e., w_{ijt}^p). In the model, we constrain these variables so that only one of them takes a non-null value for each trip type, corresponding to the promotion level selected (i.e., $p : q_{it}^p = 1$). Therefore, the overall expected revenue can be obtained by multiplying the decision variables w_{ijt}^p with parameters that describe the final price for the given promotion level: $F_{ijt}(1 - V_{it}^p)$.

Mathematical model

$$\max \quad \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}}\sum_{t\in\mathcal{T}}\left(\sum_{p\in\mathcal{P}}\left(w_{ijt}^{p}\times F_{ijt}\times\left(1-V_{it}^{p}\right)\right)-y_{ijt}\times C_{ij}\right) \tag{7}$$

$$\text{s.t.} \quad x_{i0}=W_i \qquad \forall i\in\mathcal{A} \tag{8}$$

$$x_{it}=x_{i,t-1}+R_{it}+O_{it}+\sum_{p\in\mathcal{P}}\left(\sum_{a\in\mathcal{A}}\sum_{\substack{t'=0:\\ t'+K_{a,i,t'}=t-1}}^{t-1} w_{a,i,t'}^{p}-\sum_{j\in\mathcal{A}}w_{i,j,t-1}^{p}\right)$$
$$+\sum_{a\in\mathcal{A}}\sum_{\substack{t'=0:\\ t'+DT_{a,i,t'}=t-1}}^{t-1} y_{a,i,t'}-\sum_{j\in\mathcal{A}}y_{i,j,t-1} \qquad \forall i\in\mathcal{A},t\in\mathcal{T}\setminus\{0\} \tag{9}$$

$$\sum_{p\in\mathcal{P}}\sum_{j\in\mathcal{A}}w_{i,j,t}^{p}+\sum_{j\in\mathcal{A}}y_{i,j,t}\le x_{it} \qquad \forall i\in\mathcal{A},t\in\mathcal{T} \tag{10}$$

$$w_{ijt}^{p}\le r_{ijt}\cdot D_{ijt} \qquad \forall p\in\mathcal{P},i\in\mathcal{A},j\in\mathcal{A},t\in\mathcal{T} \tag{11}$$

$$w_{ijt}^{p}\le q_{it}^{p}\cdot D_{ijt} \qquad \forall p\in\mathcal{P},i\in\mathcal{A},j\in\mathcal{A},t\in\mathcal{T} \tag{12}$$

$$\sum_{p\in\mathcal{P}}q_{it}^{p}=1 \qquad \forall i\in\mathcal{A},t\in\mathcal{T} \tag{13}$$

$$r_{ijt}=\sum_{p\in\mathcal{P}}Z_{ijt}^{p}\cdot q_{it}^{p} \qquad \forall i\in\mathcal{A},j\in\mathcal{A},t\in\mathcal{T} \tag{14}$$

$$q_{it}^{p}\in\{0,1\} \qquad \forall p\in\mathcal{P},i\in\mathcal{A},t\in\mathcal{T}$$
$$y_{ijt}\in\mathbb{Z}_0^+ \qquad \forall i\in\mathcal{A},j\in\mathcal{A}:j\neq i,t\in\mathcal{T}$$
$$r_{ijt}\in\mathbb{Z}_0^+ \qquad \forall i\in\mathcal{A},j\in\mathcal{A},t\in\mathcal{T}$$
$$x_{it}\in\mathbb{Z}_0^+ \qquad \forall i\in\mathcal{A},t\in\mathcal{T}$$
$$w_{ijt}^{p}\in\mathbb{Z}_0^+ \qquad \forall p\in\mathcal{P},i\in\mathcal{A},j\in\mathcal{A},t\in\mathcal{T} \tag{15}$$

The objective function (7) maximizes the contribution margin of the carsharing system. It has two main components: the revenue from the fulfilled trip requests, considering their final price, and the cost of transferring cars to meet demand (relocations). The final price of the trips depends on the base price per period and duration (known) and the promotion offered (a decision of the model). The final price is multiplied by the decision variable that represents the number of trips of a given type that are fulfilled for a given promotion level to obtain the total revenue. As described before, this decision variable can only take a non-null value for the promotion level that is offered (see Constraints (12)). Maximizing the contribution margin is a fairly general approach: Jorge et al. (2014), Huang et al. (2021), Liu et al. (2022), as examples, consider the revenue from rentals and the cost from relocations in the objective function, often in combination with other sources of cost that are relevant for the problems studied in these works. This allows the translation of relevant practical aspects of operations, where cost-inducing actions (e.g., relocations) are taken when they lead to additional gains in the overall expected revenue. In Section 4.3, we compare the performance of the model which uses the proposed objective function to a simpler one with a concave revenue function. We observe that both models perform similarly in terms of computational times, while the proposed model leads to decisions that are better aligned with environmental considerations as it leads to a reduced number of operator-based relocations, and consequently to reduced carbon emissions.

Constraints (8) define the initial distribution of the fleet through different areas. Constraints (9) track the fleet levels at each location and period, considering the previous number of cars in that location, the expected arrivals from relocations and trips from previous planning horizons (parameters to this model), the cars that arrive and those that left in user trips (decisions), and those arriving and departing in relocation transfers (decisions). Then, Constraints (10) ensure that the number of user trips and relocations starting at a given area and period is limited by the stock of available vehicles at that point.

Additionally, the number of fulfilled trips is limited by the demand for a given trip, whose limit is set by the expected share of the carsharing option, i.e., the probability of the representative user choosing carsharing times the total potential demand for that trip, as defined by Constraints (11). Constraints (12) ensure that the decision variable representing the number of fulfilled trips for a given promotion level can only be greater than zero if that promotion level is selected for that specific trip. Constraints (13) ensure that one and only one promotion level is selected for each combination of pick-up area and pick-up period. Constraints (14) define the probability of the representative user choosing carsharing for a given trip, depending on the promotion level offered. It should be noted that Z_{ijt}^{p} is a parameter pre-calculated for every promotion level and trip type (see Eq. (5)) based on the logit choice model and the utility function presented before. Finally, Constraints (15) represent the domain of the decision variables.

3.3. Limitations of the proposed models

Non-strategic agents:. We assume that users of the carsharing system arrive at the platform with the intention of booking one car for a trip type. We do not consider the possibility of iterative decision-making, where users compare the utility value at different times to optimize their travel decisions. We assume the users need to take the trip at a specific time and therefore do not change the time of their trip by exploring time-dependent prices, which is a common assumption in dynamic pricing literature. In reality, user behavior may indeed be strategic and involve multiple comparisons over time, nevertheless, this is more frequent on occasional and more prolonged trips, such as those seen in the car rental market. We generalize this case, which is common for typical urban mobility trips, for all carsharing users in this system. Nevertheless, extending our model to capture more nuanced user behavior such as comparing prices for different starting times and nearby locations before making a purchasing decision would be interesting and would increase our model's applicability to other travel settings.

Homogeneous users assumption:. The model assumes homogeneity among users, not considering socio-demographic characteristics that can influence mobility needs and preferences. The model might overlook important variations in demand patterns by not incorporating factors such as age, income, or travel habits. This assumption could limit the model's applicability in diverse urban environments where user behavior may be more heterogeneous. In this case, we can use a more complex choice model that accounts for the existence of latent classes among the customer population. Integrating a different choice model into our framework is relatively straightforward. Nevertheless, replacing the simple binary logit model with a more complex model, such as the latent class logit choice model or the more general mixed logit model, would require more data points and longer computational time when estimating the model parameters. We can train an appropriate choice model by considering the trade-off between prediction accuracy and computational time as typical in other applications of DCMs in traffic, if a more accurate demand model is desired. To showcase this possibility, we have applied both the binary logit and a latent class model to a real dataset from Milan. The details of the choice model and the results are provided in Section 4.4.

Binary choice assumption:. The use of a binary logit model defines the user's choice as choosing the carsharing service or opting for an outside alternative, aggregating potential different transportation options in the latter. While this modeling choice does not fully capture the complexity of real-world decision-making where users might consider multiple modes of transport, it aids in model tractability and provides a first approach to this problem. Future extensions of the model could incorporate a multinomial choice framework, allowing for a more comprehensive analysis of user preferences and behaviors.

Time to travel assumption:. The model treats travel time as a deterministic parameter, assuming we know the trip duration depending on where and when it starts and where it ends, which may not reflect the variability experienced in real-world conditions. Factors such as traffic congestion, road closures, and varying speeds can introduce significant uncertainty in travel times. This assumption can be removed by employing appropriate techniques from stochastic or robust optimization literature.

Known total potential demand assumption:. We employed this assumption for the sake of simplicity. There are several ways to relax this in future work. For example, we can model the arrival of customers as a stochastic process, estimate its parameters using historical data and study the operational problem in this setting. Alternatively, we can employ stochastic or robust programming techniques to incorporate the uncertainty in our model directly. In all cases, studying the problem under the fixed potential demand assumption first, as we have done so here, is relevant and essential.

We have used a relatively simple model to characterize the deterministic component of each alternative that included price, location of the demand, and the starting time of the trip. There are several other factors, such as convenience and comfort level, that might affect the choice. Often these attributes are qualitative and hard to measure. Therefore, finding reliable datasets quantifying these is a challenge. The operator might employ discrete choice experiments or focus groups, if they wish to include these in the demand model. Below, we work with an empirical dataset that was available to us that included a limited number of attributes. While this might be a limitation of our empirical observations, the overall framework discussed here is general and free of this limitation. Including more attributes into binary logit or any other parametric choice model, and even using a fully data-driven classification model, is possible and relatively straightforward. We discuss the use of a latent class logit model in Section 4 as an example.

4. Computational experiments

This section starts with a full description of the data and market simulation framework used in our analyses, and a summary of the experiment's environment and settings (Section 4.1). Then, the results are analyzed following two perspectives: Section 4.2 focuses on the demand models and discusses their performance and Section 4.3 focuses on the results that are relevant to the carsharing application and related insights. Then, in Section 4.4 the proposed approach is applied to carsharing trip data from Milan, Italy. Finally, Section 4.5 discusses the main managerial insights derived from this work.

4.1. Data, simulation framework, and test environment

This section describes the instances developed to test the proposed models, as well as the algorithms that simulate the responses of the carsharing market to the pricing and operational decisions. Finally, the test environment is described.

4.1.1. Instances

The data used to validate this approach was randomly generated, considering realistic values and assumptions for one-way carsharing systems. For the time-space model, we discretized the time horizon of one week into 2-hour slots, resulting in 84 periods, and considered 4 different areas within the carsharing operating system. For the potential demand pattern across time and areas, we conceptualized typical patterns in this type of mobility and randomly generated potential demand for each time period following a normal distribution with a peak around the middle of the half day. The potential demand in each period was distributed among the pick-up and drop-off areas according to a predefined distribution among the areas, characterizing them in terms of their attractiveness for pick-up and drop-off (which may be different). The total fleet was set at 10% of the average total potential demand of a half day. The time to travel between two areas depends on the origin and destination areas, as well as the start time of the trip. If the origin and destination areas are the same, this time is considered to be one period, otherwise it can vary between 1 and 3, depending on the pick-up time. The relocation costs for different origin and destination areas were randomly generated between 1 and 6. These costs consist mainly of the staff cost for driving a vehicle. Please note that the relocation costs are fixed for each pair origin/destination since the time to travel between them is also previously defined. As for prices, we assumed all base prices per period to be 3, with one price step. The promotion levels are equally distributed and their range can go up to 20% discount (in a setting of regular promotions) or 100% discount (extended promotions). This full dataset and instances are available at the following link: https://drive.inesctec.pt/s/MbbX9JZKnZwEYEW. In summary, although the demand patterns are based on typical patterns, most of the features are randomly generated (as is often the case in this literature). This allows for a solid methodological analysis of the proposed approaches in a controlled setting.

4.1.2. Market simulator

As shown in Fig. 1, to validate and assess the proposed binary logit model and the advantages and limitations of the MILP model for pricing and relocation, we simulate the market response to the promotions offered and the resulting final pricing conditions. Although the results will depend on the simulated data, the proposed market simulator is crucial to validate the theoretical advantages of this approach to support an empirical implementation of this type of learning system. The goal of the market simulator is to provide two types of outputs: (i) the observed demand after the optimized promotion levels (for the estimated demand) are made public to the users (i.e., with the price as an input), and (ii) the consequently updated values of other system states (regarding, e.g., fleet levels in different areas). In other words, this component first simulates how the market responds to the prices offered by the operator (i), then also calculates the consequences of the resulting demand in the system (ii). To obtain the simulated demand (in the face of the decided prices, which are fixed inputs to the simulation), we assume that the utility of each user for a given trip depends on the deterministic part of the utility and an error term, as shown in Eq. (1), where we assume that the set of parameters of the deterministic part is known by the market simulator. This set is different for each trip type (i.e., each combination of pick-up area i and period t and drop-off area j; not considering the categories introduced for one-hot encoding). Eq. (16) represents the deterministic part of the utility function computed by the market simulator. Comparing Eq. (16) with Eq. (6), the parameters represented by $'$ are those that control the (simulated) market responses (unknown to the remainder of the system).

$$z'^{p}_{ijt} = \alpha' + \beta'^{1}_{t} + \beta'^{2}_{i} + \beta'^{3}_{j} + \beta'^{4} \cdot F_{ijt}\left(1 - V^{p}_{it}\right) \tag{16}$$

Algorithm 1 Algorithm for the market simulation of a given trip type (combination of pick-up area i and time period t and drop-off area j) and given promotion level p such that $q^{p}_{it} = 1$.

Require: $D = D_{ijt}$ ▷ Potential demand for the trip
Require: α', $\boldsymbol{\beta'}$ ▷ "Real/ground truth" parameters
Require: $\Psi()$ ▷ Error probability distribution
Require: price ▷ Final price charged (price= $F_{ijt}\left(1 - V^{p}_{it}\right)$)
  $Q()$ ← Quantile function of $\Psi()$
  RealDemand = 0
  for $n = \{1, \ldots, D\}$ **do**
    $z' = \alpha' + \beta'^{1} + \beta'^{2} + \beta'^{3} + \beta'^{4} \cdot$ price
    z' ← standardization (z')
    x ← Random number in $[0, 1]$
    utility ← $z' + Q(x)$
    if utility> 0 **then**
      Real Demand = Real Demand +1
    end if
  end for
  return Real Demand

Algorithm 1 presents the pseudo-code for this simulation. To obtain the market response demand, the simulator calculates the utility each user attributes to carsharing. For this, the simulator receives as inputs: the potential demand for a given trip type and the parameters α', β'^{1}_{t}, β'^{2}_{i}, β'^{3}_{j} and β'^{4} (which control the actual market response), a probability distribution to describe the stochastic error term, and the final price (resulting directly from the promotion levels prescribed by the MILP model). This algorithm is run independently for each trip type. The potential demand represents the total number of users who might be interested in selecting

Table 2
Numerical example for a given trip, with potential demand of $D = 5$, resulting in carsharing 'realized' demand of 2.

n (User)	z'	Std (z')	x	Noise: $Q(x)$	Utility	User choice
1	−0.92	−0.50	0.10	−0.83	−1.33	Outside alternative
2	−0.92	−0.50	0.02	−1.35	−1.85	Outside alternative
3	−0.92	−0.50	0.85	1.85	1.35	**Carsharing**
4	−0.92	−0.50	0.56	0.55	0.05	**Carsharing**
5	−0.92	−0.50	0.40	0.10	−0.40	Outside alternative

carsharing for this trip type. For each user in the pool of potential demand, the simulator calculates the standardized deterministic part of the utility and adds a random noise to it. The random noise value is obtained from the probability distribution received as input. We consider the outside option utility to be zero. Therefore, if the overall carsharing utility is greater than zero, we consider this specific user to be demanding the carsharing service, thus increasing the simulated real demand by one. This process is repeated for all users that compose the trip's potential demand.

Table 2 presents a small numerical example to illustrate the processes in Algorithm 1. This example considers a given trip type (starting in area i at period t and ending in area j) with $D = 5$ potential demand. The error is Gumbel-distributed, and the price charged for this trip is 3.4. The parameters of the utility function are $\alpha' = 1.75, \beta'^1 = 1.02, \beta'^2 = 2.05, \beta'^3 = 1.74, \beta'^4 = -2, 2$. For each user $n = \{1, \dots, 5\}$, we calculate z', according to Eq. (16), which will be the same for all users: $z' = 1.75+1.02+2.05+1.74-2.2\times3.4 = -0.92$. We apply the commonly used z-score standardization method for its ability to transform the data to a standard scale without altering the shape of its distribution, making it suitable for algorithms sensitive to varying scales. Additionally, z-score standardization simplifies interpretation by centering the data around zero with a standard deviation of one, aiding in easier comparison and analysis across variables. To calculate the mean and standard deviation, we calculate z' for all trip types, resulting in a mean value of -0.87 and a standard deviation of 0.1. For this specific trip type, we standardize z', resulting in $\frac{-0.92-(-0.87)}{0.1} = -0.50$. Then, a random number between 0 and 1 x is generated for each user (fourth column). The quantile function of the standard Gumbel distribution leads to the noise considered. For example, for $n = 1$, $Q(0.10) = -\log(-\log(0.10)) = -0.83$. The final utility (sixth column) is obtained by adding the noise to the deterministic standardized utility. For $n = 1$, it leads to $-0.50 + (-0.83) = -1.33$. Since the outside alternative is set to have a null utility, the user chooses carsharing if the final utility is greater than zero. Therefore, user $n = 1$ chooses the outside alternative. Overall, this procedure results in a demand of two users (out of five).

The market simulator is run for all trips in the instance. The whole modeling framework, consisting of parameter estimation, MILP model run, and market simulation (see Fig. 1) is run for the full time horizon and then repeats for a given number of iterations (in a rolling-horizon approach), as will be described in the following sections. This aims to represent horizons where the demand function inputs are the same (e.g. weeks), so at each iteration, learning is accumulated.

4.1.3. Test environment setting

Experiments were run on a personal computer with an Intel Core i7 processor (i7-8550U) running at 1.80 GHz with 16 GB of installed RAM, running 64-bit Windows 10 Pro. All algorithms (except for the binary logit model estimation) were coded in C++. The mathematical programming solver used was CPLEX 12.8. The binary logit estimation model was run using the Biogeme Python package (Bierlaire, 2023). Biogeme is a software package distributed under an open-source license designed to estimate discrete choice models and is widely used in transportation research. It supports a range of discrete choice models, including binary logit models, allowing users to detail specific utility functions incorporating various attributes and estimate complex models.

4.2. Analyzing the demand estimation model

The MILP model for pricing and relocation proposed in this work embeds a binary logit model to estimate demand, which is based on an innovative utility function that incorporates both the price and the spatial and temporal features of trips. The goal of Section 4.2 is to analyze this utility function (or demand model) independently of the MILP under different settings and configurations. To evaluate the performance of the demand estimation model, this section presents a comprehensive experimental setting based on an in-sample approach (described in Section 4.2.1). Then, the results are presented and discussed in detail (Section 4.2.2), followed by an out-sample validation of the overall results (Section 4.2.3).

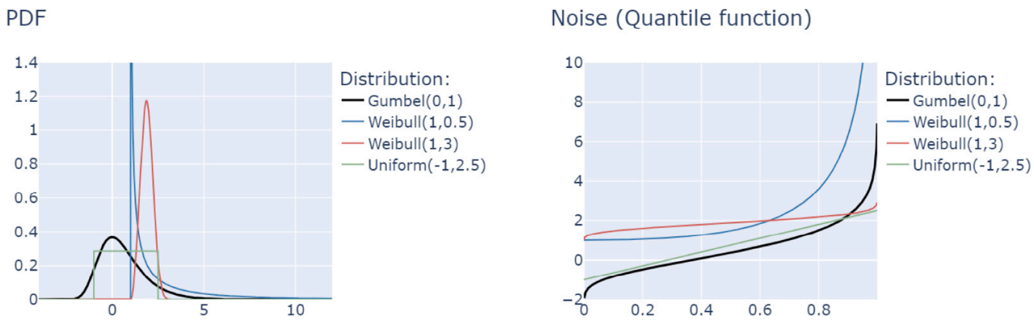
4.2.1. Experiments

For a preliminary validation of the binary logit model proposed in Section 3.1, as well as of the categorizing approach that allows accelerating the estimation procedure, we followed an in-sample approach structured as follows.

Types of categorizing strategies:. The categories for areas and periods (presented in Section 3) aim to represent the empirical knowledge of the carsharing operator regarding different peak and off-peak patterns, as well as high and low demand areas. To emulate this empirical knowledge and assign each area $i \in \mathcal{A}$ and period $t \in \mathcal{T}$ to a category, we applied an approach based on the utility parameters fed to the market simulator, which are specific to each area and time (see Section 4.1). More specially, we aim to group similar data points together, using the utility that users give to them as criteria for similarity. This means that we mean to group time periods that have a similar effect on users' proneness to using carsharing, as well as areas that have a similar attractiveness to users as pick-up or drop-off points.

Table 3
Types of categorizing strategies analyzed.

Categorizing strategy type	Number of area categories	Number of time period categories	Number of estimated parameters
B1	2	2	5
B2	2	5	8
B3	2	10	13
B4	2	21	24
B5	2	42	45



(a) Probability distribution (b) Quantile distribution

Fig. 3. Distributions tested for the error term.

Since time periods are associated with a single parameter (β'^1), we follow a standard approach for one-dimensional partitioning, Jenks Natural Breaks (Khamis et al., 2018). This is a one-dimensional partitioning method that aims to minimize variance within groups and maximize variance between them. It iteratively places breaks (or boundaries) between adjacent data points to create distinct clusters (or categories), optimizing for homogeneity within categories and heterogeneity between them. To test different category structures, we enforce a number of classes corresponding to the number of time categories presented in Table 3. The considered dataset contained 84 time periods. To obtain the number of period categories to test, we progressively divided this number by two, rounding down as necessary. As the number of categories increases, the number of specific periods in each category decreases, as expected. However, it should be noted that due to the method chosen, there is always at least one period in each category. The assignment of each trip to the different types of categories is detailed in the available instances.

To build the categories that aggregate different areas, we used K-means clustering, since these are two-dimensional features (i.e., they are classified according to their pick-up and drop-off parameters, β'^2 and β'^3). K-means clustering is a widely used technique for multidimensional partitioning. It divides the dataset into K clusters (or categories) by iteratively assigning data points to the nearest cluster centroid and then updating the centroids based on the mean of the points in each cluster. This process continues until the centroids stabilize, also optimizing for homogeneity within categories and heterogeneity between them. The number of area categories shown in Table 3 was also enforced by setting the K parameter of the K-means clustering technique. Since this dataset included four different areas, we only analyzed models with two area categories, varying the number of time period categories. The full details of the resulting categories for the different tests presented in Table 3 are reported in Appendix B.

Error term assumption:. The demand model used in this work (binary logit) assumes that the probability distribution of the error term follows a Gumbel distribution (Train, 2003). Under this condition, the estimation procedure is expected to perform well. Nonetheless, it is important to validate that this assumption does not limit the model's applicability. To analyze the impact of the error assumption on the estimation performance, we run tests feeding the market simulator (see Section 4.1) with five different probability distributions, whose shapes are shown in Fig. 3. The standard Gumbel distribution is the benchmark distribution, the assumption behind binary logit models. We then selected two different Weibull distributions: one that is monotonically decreasing and another that approximates a normal distribution with a non-zero location or mean. We tested a uniform distribution to obtain a significantly different shape than the Gumbel distribution. The range of the uniform distribution is chosen to match that of a Gumbel to an extent.

Other assumptions:. We assume that the promotions to be offered can fall within five equally spaced levels, where the first corresponds to no discount and the last corresponds to a 20% discount (regular promotions). Finally, as mentioned before, we assume zero utility for the opt-out alternative, so that the carsharing utility represents a relative value.

Methodological approach:. For the in-sample tests, we generate a dataset of trips using the market simulator approach for all trip types in the instances (see Section 4.1), once for each of the regular promotion levels. Thus, these tests are based on a data set that simulates the choices of 350,000 users (5 different promotion levels for 1344 trip types, each with its different level of potential

Table 4
Average accuracy for all trip types, measured as the average number of correct predictions under the binary logit model over the total number of predictions.

	Accuracy			
Categorizing strategy	Gumbel (0,1)	Weibull (1, 0.5)	Weibull (1, 3)	Uniform (−1, 2.5)
B1	80.8%	80.5%	81.1%	73.7%
B2	80.9%	80.6%	81.2%	73.8%
B3	80.9%	80.6%	81.2%	73.8%
B4	80.9%	80.6%	81.2%	73.8%
B5	80.9%	80.6%	81.2%	73.7%
avg	**80.9%**	**80.5%**	**81.2%**	**73.7%**

Table 5
R^2 metric for all trips, quantifying the correlation between the expected market share under the binary logit model and the real observed market share.

	R^2 market share				**R^2 market share ($D \geq 4$)**			
Categ.	G(0, 1)	W(1, 0.5)	W(1, 3)	U(−1, 2.5)	G(0, 1)	W(1, 0.5)	W(1, 3)	U(−1, 2.5)
B1	17.9%	16.8%	18.7%	−32.6%	65.3%	63.7%	66.4%	23.2%
B2	18.2%	17.1%	19.2%	−32.2%	65.9%	64.2%	66.9%	24.2%
B3	18.1%	17.2%	19.4%	−32.3%	65.9%	64.3%	67.1%	24.1%
B4	18.1%	17.2%	19.3%	−32.3%	65.9%	64.3%	67.1%	24.0%
B5	18.2%	17.2%	19.3%	−32.4%	65.8%	64.1%	67.1%	23.9%
avg	**18.1%**	**17.1%**	**19.2%**	**−32.4%**	**65.8%**	**64.1%**	**66.9%**	**23.9%**

G — Gumbel, W — Weibull, U — Uniform.

demand). We consider the types of categorizing strategies presented in Table 3 and apply different probability distributions (Fig. 3) to generate the final utility of each user. The parameters of the utility function are estimated based on the simulated choices in the dataset using maximum likelihood estimation for binary logit model. We structure the following section according to different questions we will answer to validate the model's performance.

4.2.2. Results and discussion

Tables 4 and 5 summarize two key performance metrics of the proposed estimation model: accuracy and R^2 for market share estimation. We analyze the key findings according to relevant guiding questions.

Is the estimation model sufficiently accurate given different error distributions?

Regarding accuracy, for each trip type (i.e., a trip with a given pick-up area i and period t and drop-off area j), accuracy is calculated as the number of correct predictions under the binary logit model over the total number of predictions; i.e., $\frac{D-|R-E|}{D}$, where D is the potential demand for a given trip type, R is the observed demand (given by the market simulator), and E is the expected demand (calculated by the binary logit model). Table 4 shows that the average accuracy is satisfactory for the assumed Gumbel distributed error (about 81%).

For different distributions, the average accuracy shows some interesting results. Since the binary logit model is based on the assumption of Gumbel-distributed errors, it was expected that the performance of the estimation model would be higher for a dataset where user choices actually follow a Gumbel-distributed error. However, the average accuracy for some other distributions (namely, both Weibull distributions) is surprisingly good compared to the Gumbel distribution. For the uniform distribution, the performance drops, possibly due to the shape of the distribution. However, it is interesting to note that even a uniform distribution (when the parameters are fitted to the Gumbel distribution) can perform with 73.7% accuracy, suggesting that shape is not the only relevant element. Overall, these results demonstrate that the binary logit model performs well when the underlying distribution is similar to a Gumbel. This is as expected. When the distribution has a significantly different shape, the accuracy of the demand distribution decreases, but not too dramatically.

Is the estimation model sufficiently accurate for more aggregated (and thus faster) categorizing strategies?

Table 4 also shows that, interestingly, the number of categories (and therefore the number of trip types in each category) does not seem to affect accuracy significantly. This suggests that although there are a large number of areas and time intervals in the original model, several of these share common characteristics and they can be aggregated into a small number of meaningful clusters when estimating demand. Since, as expected, run time increases with the number of parameters to be estimated, using a reduced number of categories seems to be a fast and efficient strategy for the estimation procedure.

Does the model accuracy translate into ability to predict market share?

The R^2 metric (Table 5) aims to quantify the degree to which the binary logit model correctly estimates the market share of carsharing. The market share for each trip type represents the fraction of potential demand that chooses carsharing. Therefore, this metric represents the R^2 value of a linear regression between the expected market share ($\frac{E}{D}$) and the observed market share ($\frac{R}{D}$) for all trip types, calculated according to Eq. (17):

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{(E/D - R/D)^2}{(R/D - \overline{R/D})^2} \tag{17}$$

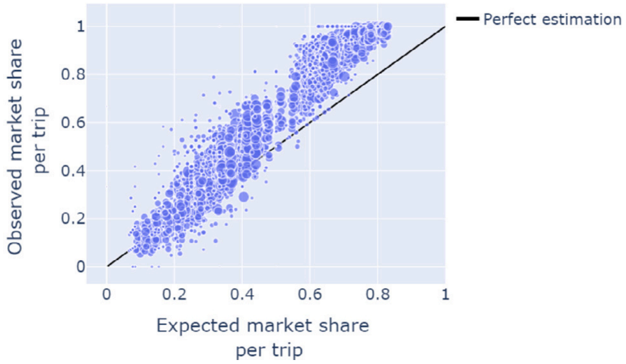


Fig. 4. Bubble chart for expected and observed market share (market share plots), for categorizing strategy B1 with Gumbel-distributed error. Each circle represents a type of trip (combination of pick-up area i and time t and drop-off area j), and the circle's width represents the size of the potential demand for that trip type. The solid line (in gray) represents a theoretical perfect estimation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Performance metrics to compare the categorizing approach with the full estimation for the truncated dataset with Gumbel-distributed errors.

Approach	Accuracy	R^2 market share	R^2 market share ($D \geq 4$)
Categorizing (B2)	80.6%	23.7%	68.0%
Full estimation	80.6%	23.4%	68.0%

where $\overline{R/D}$ represents the average of the observed market shares. In terms of trends across categorizing strategies and error probability distributions, the results for this metric seem to support the conclusions derived from the accuracy metric. However, the correlation seems to be small even for the Gumbel distributed error (18.1%). This is explained by the impact of some trip types with very low potential demand, where small absolute deviations in the number of users choosing carsharing translate into significant changes in market share. Fig. 4 illustrates this effect by showing the trip types as circles whose width is determined by the size of their potential demand. It can be seen that several smaller trip types influence the correlation line. This figure also allows us to understand that there is a systematic deviation from what would be a perfect estimate, especially for trips with a higher market share, which seem to be systematically underestimated.

Is the impact of small trip types (with very few users) significant?

To address this issue, Table 5 also shows the values of this correlation when it is calculated considering only trip types whose potential demand is more than or equal to 4 users. This corresponds to removing only 0.2% of the potential trips from the correlation calculation. In this case, the correlation values are substantially better. This suggests that we do not yet have enough information from the observations in our dataset for a small proportion of trips. If this was a real world application and there were an opportunity to conduct experiments, gathering more data related to these types of trips would potentially have improved the model accuracy.

Does the categorizing approach have a significant impact on model performance (vs. a non-categorizing approach)?

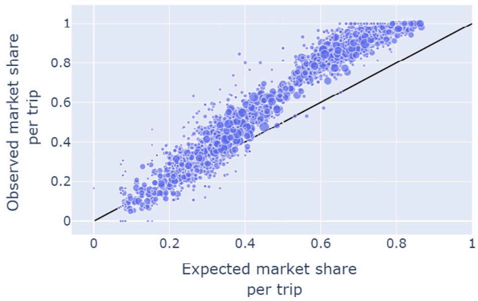
The results above show that the categorizing strategies perform similarly. However, it is important to understand how the categorizing strategies compare to a strategy where the areas and time periods are not aggregated into categories. This requires the estimation of domain and period specific parameters for each origin and destination area, and time period. The software used (*Biogeme*) could not handle this dataset with so many parameters due to the high number of time periods (84). Therefore, in order to have a theoretical exercise to analyze the performance of the categorizing approach, we truncated the dataset to consider the first half of the time horizon, thus reducing the number of time periods from 84 to 42. We analyzed the results for Gumbel distributed errors and presented them in Fig. 5. Table 6 also shows that the performance of the two approaches is similar. This supports the value of the categorizing approach, which allows simplifying and speeding up the estimation models with similar performance.

Does the trip-aggregating approach have a significant impact on model performance (vs. estimating the trips individually)?

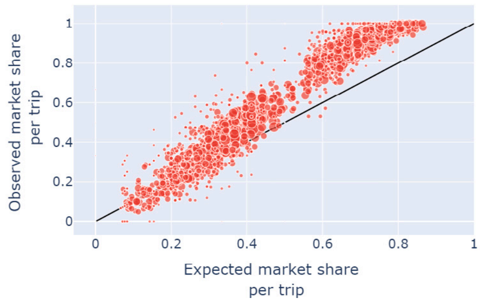
The good performance of the abovementioned approaches seems to be due to the integrated overview of carsharing trips. That is, instead of considering and estimating each trip type independently, we aggregate all trip types under the same utility function, considering the pick-up and drop-off as features. This allows the information from other trip types that share the pick-up area, for example, to be used to estimate all trips. The downside is the increased complexity of the utility function, which is mitigated by the proposed categorizing strategies. As discussed in Section 2, the standard approaches to estimation focus on each trip independently and do not consider an integrated overview of the entire system in the same utility function. Comparing the aggregated estimation approach presented earlier with a standard independent approach where the price is the feature considered for each trip allows for quantifying the gains of the proposed approach. Table 7 presents the performance metrics for the independent trip estimation approach compared to the proposed aggregated approach. These results demonstrate the value of using an integrated approach rather than estimating demand for each trip independently, across all performance metrics.

Table 7
Performance metrics to compare the aggregated approach with the independent trip estimation for the full dataset with Gumbel-distributed errors.

Approach	Accuracy	R^2 market share	R^2 market share ($D \geq 4$)
Aggregated estimation (Categorizing - B2)	80.8%	17.9%	65.3%
Independent trip estimation	76.1%	−4.7%	31.4%

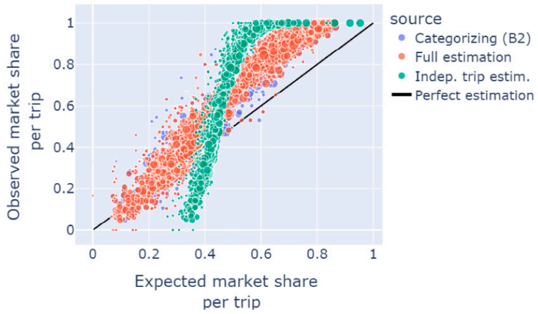


(a) Categorizing approach (B2)



(b) Full estimation

Fig. 5. Market share plots for categorizing approach vs full estimation for area- and period-specific parameters. Analysis for the truncated dataset under Gumbel-distributed errors. As in previous plots, the full black line represents perfect estimation.



(a) Market share plots

(b) Accuracy per observed market share

Fig. 6. Comparison of aggregated estimation (categorizing approach B2 and full estimation) and independent trip estimation, for the truncated dataset under Gumbel-distributed errors.

Fig. 6 attempts to show this difference. Fig. 6(a) extends the market share plot in Fig. 5 to include the independent trip estimation for the truncated dataset. It is possible to understand that there is a systematic deviation from perfect estimation. These results can be seen from a different perspective in Fig. 6(b). It is possible to see that the quality of the estimation is lower for trip types with lower observed market share, it increases linearly as the observed market share increases, up to about 40% where it reaches very good values (in Fig. 6(a) the points intersect the "perfect estimation" lines). Then, as the observed market share increases, it steadily and linearly decreases to the worst performance values. This systematic deviation seems to be price related, as shown in Fig. 7. However, due to the price-demand relationship, higher market shares are expected to be associated with lower prices (and vice versa).

4.2.3. Out-sample validation

The estimation model has shown acceptable quality for some of the probability distributions tested that do not conform to the theoretical assumptions of binary logit models. Nevertheless, for the sake of simplicity, the remainder of the paper assumes that the error terms follow a Gumbel distribution. Moreover, since the categorizing approaches have shown good and similar performance in the in-sample tests, we will use the categorizing strategy B2 in the following, with a balanced number of parameters to estimate.

To validate the conclusions of the previous section in an out-sample approach, we divided the full dataset developed for the in-sample tests into a training set and a test set. Table 8 shows the accuracy values obtained for each set, for different weights of the training set. Here, the percentage of the dataset assigned to each subset refers to the fraction of trip types randomly assigned to each subset. The weights of these sets in terms of total trips (users) also correspond to these fractions due to the random assignment procedure. The results indicate that the binary logit model is still able to provide good quality estimates in an out-sample approach, as will be discussed in the following sections.

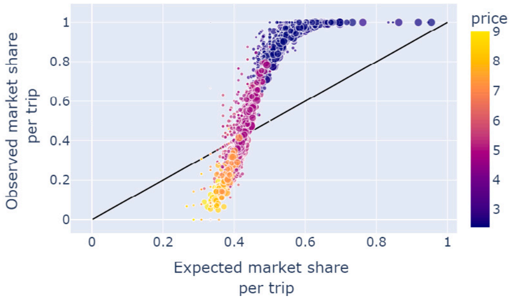


Fig. 7. Market share plots for independent trip estimation, for the truncated dataset under Gumbel-distributed errors.

Table 8
Average accuracy for training and testing set in an out-sample approach.

	Accuracy (avg)	
% trip types in training set	Training set	Testing set
20%	81.80%	80.70%
50%	81.50%	80.30%
80%	81.10%	80.09%

4.3. Analyzing the carsharing operations model

This section aims to analyze the performance of the carsharing operations model with the embedded binary logit estimation model as presented in Section 3. To do so, we follow the methodological approach in Fig. 1. The utility parameters are estimated in each iteration based on the information collected so far. Then, these parameters feed the mathematical programming model that optimizes the promotion levels to offer and the relocations to perform. The market simulator then returns the "real trips" that materialize in response to the promotion levels and fleet distribution resulting from the optimization model. These observations of demand and final prices charged accumulate with each iteration, so there is an increasing amount of data on which to run the estimation.

To start the iteration process, we generate demand for each trip type assuming no discount is offered. We run 5 iterations, with the market simulator tuned for Gumbel distributed errors and the binary logit estimator considering the categorizing strategy B2 as described in the previous section. The area and time period categories defined do not change throughout the simulation horizon. The running time of the MILP model was limited to 600 s.

To understand the impact of this methodology, we compare the results of the model proposed in Section 3 (hereafter referred to as the *binary logit and mixed integer linear program* (BL-MILP) approach) with a similar model but with a linear demand assumption (LD-MILP approach), which is often used in the literature. The LD-MILP model does not exploit the estimation capabilities of the binary logit model and does not require prior estimation of the parameters. Since the key innovation of the proposed model is the integrated learning mechanism derived from the embedded binary logit estimation model, comparing its performance with the existing approaches requires running the same model without the learning mechanism. The linear demand model applied is explained in Appendix C. In addition, to provide an upper bound on the gains from correctly estimating demand, we also run the carsharing operations model assuming perfect information about demand in each iteration (PI approach).

Finally, we present the results for two different pricing settings. The first refers to regular promotions, where the maximum allowed discount value is 20%, using the dataset and methodology described above. Then, we investigate the impact of extending the range of promotions, including the possibility of offering "free" rides (100% discount).

4.3.1. Regular promotions

Table 9 shows the results of each approach in the MILP run: the estimated profit, the run time (which was equal to the maximum time allowed for BL-MILP and LD-MILP models), and the resulting MILP gap (which was always less than 1%). In addition, Table 9 also shows the realized profit, i.e., the profit from the actual realized demand (different from the estimated demand that led to the estimated profit, except in the case of perfect information) that was observed when the promotion levels were offered and the moves were decided. Table D.1 in Appendix E details the results per iteration.

Overall, the proposed approach (BL-MILP) performs better than using a linear demand (LD-MILP) model, increasing profit by an average of 2.2%. It should be noted, however, that this is a conservative comparison. We use complete information to calculate the average demand values that support the LD-MILP model, giving it an advantage that is unlikely to be matched in reality. As will be discussed below, the proposed approach shows the potential to improve relative performance in more extended pricing settings. As an aside, the results in the appendix also show that, as expected from the previous analyses, the performance of the proposed model is relatively constant across weeks. One of the reasons that could explain this behavior is the constancy of the final prices resulting from the optimization model.

Table 9
Results for the MILP model and realized profit per iteration and per approach, considering regular promotions.

	BL-MILP	LD-MILP	PI
Estimated profit	62,057	64,625	–
Time to run (sec)	600	600	5
MILP gap	0.06%	0.32%	0.00%
Realized profit	58,703	57,427	76,598
Profit improvement	–	+2.2%	−23.4%

Table 10
Decisions on pricing (distribution of trip types per promotion level selected) and relocations, and resulting service level metrics for regular promotions (average for all iterations).

Promotion level	BL-MILP	LD-MILP	PI
	% of trip types with promotion level		
0	41%	29%	100%
1	16%	30%	0%
2	10%	30%	0%
3	10%	11%	0%
4	23%	0%	0%
Demand	41,812	43,171	39,229
Trips served	13,082	12,958	7,456
Service level	31%	30%	19%
Relocations	952	928	86

The largest difference in performance in Table 9 concerns the PI model. The fact that the proposed BL-MILP approach is still 23.4% below in terms of profit shows the importance of accurately understanding and estimating demand in this context. In fact, the overall pricing and shifting strategy under PI is substantially different than in an uncertain environment. Table 10 shows the resulting decisions for the three approaches (regarding pricing and transfers), as well as their impact on demand, trips served, and the resulting service level. Full knowledge of demand leads the model to decide not to offer discounts. That is, offering discounts seems to be a good strategy to deal with demand uncertainty by influencing it. When demand is fully known in advance, it is optimal to charge high prices and save capacity for high revenue trips. However, this strategy has an impact on the service level. Even though there is less demand in PI (due to the higher prices), it is also optimal for this approach to fulfill fewer trip requests, resulting in a lower service level compared to the BL-MILP and LD-MILP approaches. Also for transfers, the results shown in Table 10 demonstrate the importance of understanding demand in order to reduce these costly (and polluting) operations.

When comparing BL-MILP and LD-MILP pricing decisions in Table 10, it is interesting to observe how the proposed approach favors either full prices (the lowest promotion level) or large discounts (the highest promotion level, corresponding to a 20% discount). On the contrary, LD-MILP pricing decisions tend to be balanced among the first (lowest) promotion levels. Even though charging high prices seems to be a good strategy when demand is known in advance (as seen with PI), the proposed approach with higher discounts achieves better results than LD-MILP. To analyze this further and to understand whether the maximum discount of 20% limits the performance of the approach, we decided to extend the promotion levels offered, as mentioned above. These results are presented in Section 4.3.2, along with a full analysis of the impact on prices and relocations. Before that, we analyze in more detail the performance of the MILP model when scaling the problem and the impact of the current objective function structure with regular promotion levels.

MILP performance

Table 9 shows that stopping the solvers at 600 s results in a small MIP gap, smaller than 1%. We stopped the solvers at 600 s since the optimality gap does not improve much even if we allow the solver to run twice the time, as shown for one of the instances in Fig. 8. This shows that even though we can obtain very good solutions in a short amount of time, proving optimality requires much longer computational time. For the application at hand, a 1% optimality gap is acceptable.

To further understand how the model performs when the system scales, Table 11 presents the results of a sensitivity analysis on the fleet size, a critical parameter influencing the system's ability to meet demand and manage relocations efficiently. This analysis highlights the impact of scaling on the model's performance, showing its ability to respond under varying operational conditions. When the fleet is substantially reduced, finding the optimal solution might become more difficult, due to the existence of alternative uses for the reduced fleet. This is shown in the table in the first rows, where the fleet scale is smaller than 1. However, the model still performs adequately, consistently obtaining MIP gaps under 1%. When the fleet becomes larger, meeting the existing demand becomes easier, and finding the optimal solution becomes straightforward, even allowing the model to prove optimality in seconds (for fleet sizes that are three times or more larger than the original size).

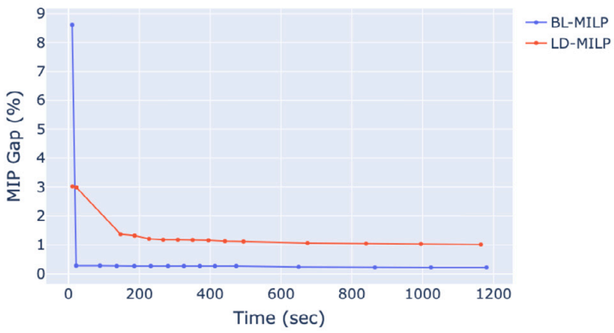


Fig. 8. Optimality-gap evolution for an instance with the BL-MILP and LD-MILP up to 1200 s.

Table 11
MILP performance under varying fleet sizes (with LD-MILP), with a time limit of 600 s. Available fleet scaled according to the original instances (represented by scale=1).

Fleet scale	Best of value	Time to run (sec)	MIP Gap
0.10	8,198	600	0.02%
0.25	18,139	600	0.03%
0.50	32,433	172	0.00%
0.75	46,184	600	0.05%
0.90	53,217	600	0.16%
1.00	57,421	600	0.35%
1.10	61,528	600	0.30%
1.25	67,432	600	0.17%
1.50	76,857	600	0.06%
2.00	93,143	600	0.03%
3.00	11,4232	27	0.00%
4.00	131,145	14	0.00%
5.00	141,075	4	0.00%
10.00	145,647	4	0.00%

Objective function analysis

In Section 3.2, we discussed that our model maximizes the contribution margin (i.e., considers revenues and variable costs). In contrast, some studies in the literature on shared mobility consider a simpler alternative and consider maximizing a concave revenue function. To understand the trade-off between these two alternatives, we solved the same instances as in Tables 9 and 10 using the simpler objective function. In this test, we used the LD-MILP approach for demand modeling in order to isolate the effects of the model from the learning component. This new run was stopped at 10 min as the others and the average MIP gap reduced slightly (from 0.32% to 0.18%), suggesting that the complexity of the model was not substantially reduced by this simplification of the objective function. In fact, as relocation decisions are still embedded in the mathematical model, this presumably results in a similar resolution process. The results are summarized in Table 12, which shows that pricing decisions are essentially similar, with slightly fewer discounts and thus slightly less demand, but with similar service levels and revenue. However, the number of (operator-based) relocations more than doubled. When considering the patterns of relocations with a revenue-maximizing objective function, they occur in every direction and at any time since there are no controls to limit the use of this balancing lever, increasing costs, both economic and environmental. While the change in economic terms might not be substantial, the significant increase in environmental consequences supports the use of the proposed objective function, even if it is slightly more complicated.

4.3.2. Extended promotions

When considering extended promotions, we consider 11 equally spaced promotion levels, where the first level ($p = 0$) corresponds to no discount and the last level ($p = 10$) corresponds to a free trip (100% discount). It should be noted that the results (such as total profit generated) for regular and extended promotions cannot be compared in absolute terms. Due to the standardization procedure explained in Section 4.1.2, the market simulator that generates demand considers the defined range of potential prices to standardize utility. Therefore, the demand response for the same discount is different in the two promotion settings. Nevertheless, the relative results, especially when comparing approaches, are valid and provide interesting insights into the problem.

Table 13 shows the overall results for each approach (BL-MILP, LD-MILP, and PI) for extended promotions as an average of all iterations (Table D.2 in Appendix E shows the detailed results per iteration). In this setting, as the MILP model grows with the addition of discrete promotion levels to choose from, the MILP gap for the 600-second time limit increases slightly for BL-MILP and LD-MILP (although it is still quite small), while it still quickly proves optimality in PI. As in the previous setting, full knowledge of demand (PI) leads to significant gains in profit. This shows that the proposed approach still has room for improvement, as it lags the PI upper bound on profit by 28.2%. Nevertheless, in this setting, the gains over the traditional LD-MILP approach are substantial (18.5%). This seems to be the impact of improved demand estimation on different decisions.

Table 12
Decisions on pricing (distribution of trip types per promotion level selected) and relocations, and resulting service level metrics for regular promotions (average for all iterations) for LD-MILP approach with different objective functions.

Promotion level	Revenue-maximizing obj. function	Obj. function considering costs
	% of trip types with promotion level	
0	29%	29%
1	36%	30%
2	29%	30%
3	6%	11%
4	0%	0%
Demand	42,795	43,171
Trips served	12,862	12,958
Service level	30%	30%
Relocations	1,955	928
Revenue	59,040	58,513

Table 13
Results for the MILP model and realized profit per iteration and approach, considering extended promotions.

	BL-MILP	LD-MILP	PI
Estimated profit	62,057	43,104	–
Time to run (sec)	600	600	20
MILP gap	0.06%	1.08%	0.00%
Realized profit	47,523	40,089	66,192
Profit improvement	–	+18.5%	−28.2%

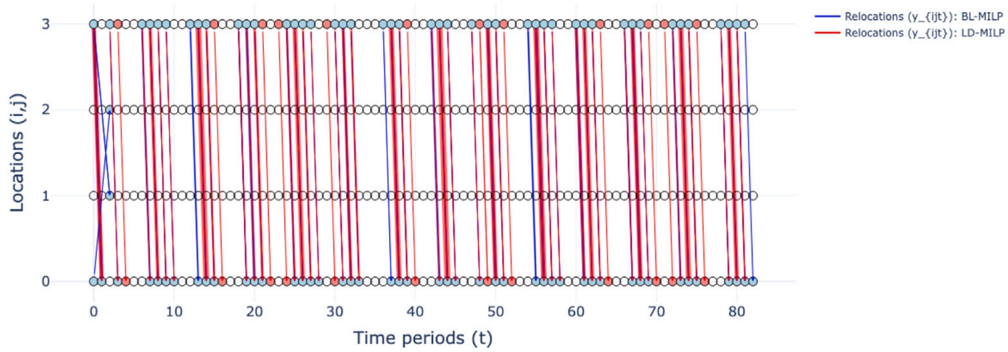


Fig. 9. Visualization of the relocation movements resulting from running one iteration of the proposed approach (BL-MILP) and the no-learning approach (LD-MILP). The arrows indicate the orientation and duration of the relocation of the cars. The width of the arrows indicates the number of relocated cars.

Looking more closely at the pricing decisions (Table 14), it is clear that with full knowledge of demand (PI), it is optimal to charge full prices for all trips. As before, this leads to a lower level of demand, but also to a much smaller proportion of trips being served, resulting in a lower level of service. As for the other approaches, discounts are often used, with BL-MILP going up to 50% discount ($p = 5$) and LD-MILP up to 60% discount ($p = 6$). Nevertheless, more than 60% of the trips in the BL-MILP approach have at most a 10% discount, while this value drops to 37% in the LD-MILP approach. The lower prices in LD-MILP lead to higher levels of demand that the system cannot (or does not find it optimal to) fully satisfy, resulting in a low level of service. These higher demand levels may help explain the high levels of relocations in LD-MILP, which may be necessary or profitable to (potentially) serve upcoming trips. In terms of relocations, this value is significantly lower for PI, as before and as expected, since with full knowledge of the demand, any relocation decided by the model is definitely helpful to serve a profitable upcoming trip. Fig. 9 shows, as an example, the relocation results for one of the iterations of the proposed approach (BL-MILP) and the no-learning approach (LD-MILP). The graph is built on a space–time network where the nodes represent combinations of locations and time periods. The different locations are represented in the vertical axis, while the horizontal axis shows the time. An arrow represents a relocation movement, both spatially and temporally. The width of the arrows is proportional to the number of cars that are relocated. The results of the model allow to know when and where the relocations start and end, as it is shown in the figure. It is possible to see that the no-learning approach resorts substantially more to relocations. Moreover, there is a clear trend for moving cars from location 3 to location 0. This shows that the former is likely a typical “drop-off” location whereas the latter is where demand is concentrated.

Interestingly, even under the assumption of perfect demand information, relocations, which have a significant environmental impact and are costly to operate, are still used. This is mainly due to the uneven pattern of carsharing demand, which is well known

Table 14
Decisions on pricing (distribution of trip types per promotion level selected) and relocations, and resulting service level metrics for extended promotions (average for all iterations)

Promotion level	BL-MILP	LD-MILP	PI
	% of trip types with promotion level		
0	44%	29%	100%
1	20%	8%	0%
2	3%	24%	0%
3	9%	21%	0%
4	16%	14%	0%
5	8%	2%	0%
6	0%	2%	0%
7	0%	0%	0%
8	0%	0%	0%
9	0%	0%	0%
10	0%	0%	0%
Demand	25,231	33,323	21,761
Trips served	13,275	12,408	6,562
Service level	53%	37%	30%
Relocations	667	1019	180

in the literature. However, the relocation cost parameter also plays a relevant role in this decision. In this context, the relocation cost was assumed to be dependent on the trip duration. However, in order to understand its full impact on the relocation and pricing strategies, we performed a sensitivity analysis on this parameter, the results of which are presented in Table 15. For the three approaches compared, we show the impact of the relocation cost (ranging from 0 – free relocations – to a very large number) on the total profit, the number of relocations, and the average promotion level (which can range from 0 to 10, where 0 represents no discount and 10 represents a free trip), as well as the changes in profit and relocations when compared to the original value of the relocation cost parameter. As expected, free relocations lead to the highest profit and number of relocations for all approaches. However, the highest number of moves for PI is significantly lower than for the other two approaches. It is also interesting to note that with full knowledge of the demand, costly relocations (100) are still optimal to perform, since they end up serving more than one trip (which can be inferred from the maximum revenue of the trips). In BL-MILP and LD-MILP, the average promotion level has an interesting non-monotonous effect as the cost of relocation increases. While performing relocations is still optimal, discounts tend to increase as relocation costs increase. This makes sense because discounts can be used as a demand-shaping mechanism to (only partially) replace relocations. When relocation becomes too costly, the average discount tends to stabilize at a slightly lower level. In summary, although relocations can be an extremely flexible tool to meet demand, their cost has a significant impact on profit and (although not directly considered in these models) on the environment. The fact that the proposed approach allows to reduce the number of relocations (compared to LD-MILP) in the extended promotions setting is a clear advantage. The fact that the PI bound shows that there is still much room for improvement in terms of relocations, which can be addressed through better demand estimation and management, is an indication that further research on this topic is worthwhile.

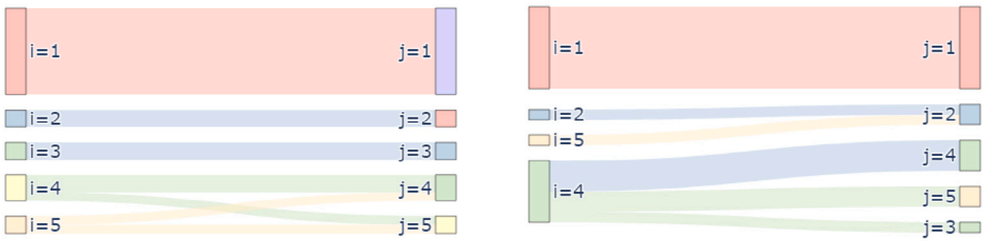
4.4. Application to a carsharing system in milan, Italy

We applied the proposed approach to data from a carsharing system in Milan, Italy, made available in Archetti et al. (2023). Due to the lack of public data on carsharing (or other shared mobility modes) regarding the prices charged per trip, we used the real dataset to represent the operational system and demand patterns while the market simulator simulated the user responses to the prices charged.

Data:. The data available consists of carsharing trips in Milan. We considered trips from one full day of operations (720 min) and 105 locations within the city center. The trips are characterized by the pick-up and drop-off locations and times. We aggregated the locations into five areas and the rolling-time pick-up datestamps into twenty time periods to build trip types with meaningful demand values. The observed demand in this dataset for each trip type was used as potential demand. We calculated the driving time between areas (depending on the start time) based on the average durations of those trips. Fig. 10 represents the demand flows between areas for different time periods for the Milan carsharing system, showing how they differ depending on the areas and throughout the day. The data also included the number of cars idle in each location at the beginning of the time horizon, as well as the number of cars currently in use, and where and when they would become available during the time horizon. This dataset did not incorporate relocated cars. We held all data related to the pricing setting as in the previous tests due to the lack of real data. Overall, it should be noted that the real dataset was substantially smaller than the simulated dataset we used to validate the approach in terms of total demand served. Nevertheless, a substantially higher ratio of cars to demand led to lower utilization and higher service level metrics, as discussed later.

Table 15
Sensitivity analysis to the relocation cost — results for profit, relocation and average promotion level. Promotion levels range from 0 (no discount) to 10 (free). Original cost: in average.

	BL-MILP				
Relocation cost	**Profit**	**Relocations**	**Avg promotion level**	**Δ profit vs original**	**Δ relocations vs original**
0	48,599	2570	1.6	2%	270%
5	45,190	75	1.8	−5%	−89%
10	43,464	50	2.0	−9%	−93%
100	40,875	0	1.9	−14%	−100%
10000	40,905	0	1.9	−14%	−100%
	LD-MILP				
Relocation cost	**Profit**	**Relocations**	**Avg promotion level**	**Δprofit vs original**	**Δ relocations vs original**
0	48,101	2550	0.8	7%	170%
5	44,038	56	1.1	−2%	−94%
10	43,782	50	1.1	−3%	−95%
100	41,280	0	1.0	−8%	−100%
10000	41,280	0	1.0	−8%	−100%
	PI				
Relocation cost	**Profit**	**Relocations**	**Avg promotion level**	**Δ profit vs original**	**Δ relocations vs original**
0	66,622	557	0	0%	259%
5	66,142	50	0	0%	−68%
10	65,892	50	0	−1%	−68%
100	61,392	50	0	−7%	−68%
10000	60,695	0	0	−9%	−100%



(a) $t = 2$ (7:36AM-8:12AM)

(b) $t = 9$ (11:48AM-12:24PM)

Fig. 10. Demand flows between areas in different time periods. The pick-up areas are represented in the left and the drop-off areas in the right.

Tests:. As in the tests with simulated data presented in the previous sections, we run five iterations of the proposed BL-MILP approach, as well as the standard LD-MILP approach (where no learning takes place) to compare the results. We estimated the parameters using Biogeme in each iteration, with increasing amounts of accumulated data. We aggregated the areas and time periods using the same categorizing method and number of categories. Additionally, two types of promotions were tested as before: regular promotions (up to 20% discount) and extended promotions (theoretically up to 100% discount).

Results:. Table 16 shows the average results per iteration for the MILP model for both approaches for regular and extended promotions. Due to the size of the dataset, the model consistently nearly immediately achieves optimality in all situations. Considering the leading indicator of realized profit, the satisfactory performance of the proposed approach versus an approach with no learning is validated, with slight improvements for regular promotions and a substantial increase when extended promotions are considered (as observed with simulated data). Table 17 details the pricing decisions that led to these results, as well as additional insights on demand levels, demand served, service level and relocations for the setting with regular promotions. When considering the decisions on pricing under regular promotions, similar to the results with simulated data, the LD-MILP approach settled on lower prices on average. Although not as clearly, a similar pattern can be observed for extended promotions (Table E.1 in the Appendix). Since the number of trips is much smaller in the new dataset, it is more challenging to draw significant conclusions. In the Milan dataset, there is a very high service level for both approaches (98% vs 30%–53% in the simulated dataset) and very few relocations (at most 36 relocation trips vs a minimum of 667 in the simulated dataset). This is due to the substantially higher ratio of cars to potential users. Therefore, it is challenging to assess the value of our approach in these dimensions in the Milan dataset.

Overall, the dataset from Milan confirms the improvements achieved with our approach, especially concerning the system's profitability. Since the dataset is small, it is challenging to extrapolate additional insights, namely related to service level or relocations. Moreover, since these analyses are based on the market simulator proposed in this paper (since real data on prices is unavailable), there is still room to extend this approach to a real-life setting, requiring real-time price experimentation and monitoring.

Table 16
Results for the MILP model and realized profit per iteration and per approach, considering regular and extended promotions, for the real dataset.

	Regular promotions		Extended promotions	
	BL-MILP	LD-MILP	BL-MILP	LD-MILP
Estimated profit	1,235	2,039	518	1,105
Time to run (sec)	1	0	1	0
MILP gap	0%	0%	0%	0%
Realized profit	1,329	1,317	688	610
Profit improvement	+1.0%		+12.7%	

Table 17
Decisions on pricing (distribution of trip types per promotion level selected) and relocations, and resulting service level metrics for regular promotions (average for all iterations) for the real dataset.

Promotion level	BL-MILP	LD-MILP
0	36%	41%
1	0%	0%
2	24%	0%
3	16%	32%
4	25%	26%
Demand	195	193
Trips served	190	189
Service level	98%	98%
Relocations	35	36

Latent class logit model:. Here we explore the adaptation of the general framework to a different choice model. We work with the latent class binary logit model which allow identifying subgroups (latent classes) within a population based on observed data. Latent class models are useful when the population is heterogeneous by clustering individuals into distinct classes based on their response patterns. The model assumes that individuals belong to one of a finite number of latent (unobserved) classes and that their responses are probabilistically determined by the class to which they belong. For our application, this means that Eq. (6) is adapted to account for class-dependent β-type parameters, and the deterministic part of the utility of a given trip type is z^k_{ijt} where $k = \{1, \dots, K\}$ is the latent class (and, consequently, for each price level p, it is given by z^{pk}_{ijt}). Therefore, Eq. (4) becomes the weighted sum of the probability of a user selecting carsharing conditional on being in class k, weighted by the class prevalence in the population of class k, denoted as π_k:

$$\text{share} = \sum_{k=1}^{K} \pi_k \frac{\exp(z^k_{ijt})}{\exp(z^k_{ijt}) + \exp(0)} \tag{18}$$

and Eq. (5) is modified to:

$$Z^{pk}_{ijt} = \sum_{k=1}^{K} \pi_k \frac{\exp(z^{pk}_{ijt})}{\exp(z^{pk}_{ijt}) + \exp(0)}. \tag{19}$$

We estimated the latent class model with 2 classes on the empirical dataset used in this section using Biogeme. The results were similar, with the latent class model slightly overperforming, as expected, with an average of 79.5% accuracy (vs 79.2% for the binary logit). The accuracy is calculated as described in Section 4.2.1, excluding trips with very few potential users (less than 4).

4.5. Managerial insights

This work provides valuable insights for carsharing operators, highlighting the importance of accurate demand modeling, strategic use of promotions, and integration of advanced mathematical models for more efficient and sustainable operations. The results suggest that the use of a binary logit model within a mixed-integer linear programming framework provides a powerful approach for optimizing carsharing systems in the face of dynamic demand and operational challenges. The key managerial insights are as follows:

Robustness of demand modeling:. The estimation model shows robustness to different error distributions. While it was designed based on Gumbel distributed errors, it also performs reasonably well for other error distributions. This highlights the model's adaptability to different scenarios, which is critical for real-world applications.

Aggregation of trip information:. Aggregating trip information using a unified utility function improves model performance. The proposed approach, which integrates different trips under a single utility model, proves superior to estimating each trip independently. This aggregation, coupled with the categorizing strategy, streamlines the estimation process without compromising accuracy.

Value of the proposed BL-MILP approach:. The BL-MILP approach consistently outperforms linear demand models, demonstrating the value of integrating a binary logit model into operational decision-making. The dynamic interplay between demand estimation and operational decisions leads to improved profitability and a more efficient supply–demand balance than traditional linear approaches. Also, the proposed approach shows robust performance in out-sample validation. This demonstrates the framework's ability to provide accurate estimates and support operations decisions even in unseen scenarios and reinforces the practicality of the approach in dynamic and evolving operational environments. These conclusions are supported by the analysis of a carsharing trip dataset from Milan, Italy.

Impact of promotions:. Regular and extended promotions have a significant impact on operational decisions and profitability. For regular promotions, the proposed approach increases profits by an average of 2.2% over the traditional linear demand model. Extended promotions (which can go up to 100%, i.e., offering free rides) reveal the substantial gains of the approach, outperforming the traditional model by 18.5%. The results demonstrate the importance of using the binary logit model for demand estimation in designing effective promotion strategies. Although the proposed approach still falls short of the perfect information upper bound by up to 28.2%, this demonstrates the potential for improved performance in uncertain environments.

Impact of relocations:. Relocation costs have a significant impact on operational decisions and environmental considerations. Sensitivity analysis shows that while relocation is a flexible tool to meet demand, its cost affects profit and environmental impact. The proposed BL-MILP approach demonstrates an advantage in reducing relocations compared to the LD-MILP model, highlighting its potential for more sustainable and cost-effective operations.

5. Conclusions

This work addresses the challenges of balancing supply and demand in free-floating one-way carsharing systems by proposing a mixed-integer linear programming model incorporating a binary logit model for demand estimation. Unlike traditional approaches, this integrated model uses the binary logit to both estimate and manage demand efficiently. Innovatively, the utility model uses the pick-up time, pick-up area, and drop-off area as features of the trip, besides price, which allows the integration of critical information related to spatial and temporal properties. This information can be acquired from different trip types, leading to enhanced demand learning. A categorizing strategy is introduced to improve computational efficiency in the face of the number of features considered. The effectiveness of the proposed approach is validated through extensive computational experiments and sensitivity analyses on simulated data, as well as actual carsharing trip data from Milan.

This paper's findings provide actionable insights for carsharing operators and underscore the importance of accurate demand modeling and dynamic decision-making in achieving sustainable and profitable operations. The proposed model's adaptability to different error distributions underscore its effectiveness in dynamic operational environments. The integrated overview of all trips (instead of considering each trip independently) brings substantial advantages to the estimation process due to the use of trip characteristics as features besides price. The model's effectiveness is not only evident in regular promotions, but becomes particularly pronounced in extended promotions, where the proposed approach consistently outperforms traditional linear demand models. The increased performance of the proposed approach is validated in this dimension by the analysis of a carsharing trip dataset from Milan, Italy. Moreover, sensitivity analysis underscores the importance of relocation costs, revealing their significant impact on profit and environmental considerations. The proposed approach demonstrates an advantage in reducing relocations compared to traditional models, consistent with sustainability goals and cost-effectiveness.

In conclusion, this study contributes not only methodologically but also practically to the field of carsharing operations. The findings highlight the critical role of accurate demand modeling and dynamic decision-making in designing effective promotions and ensuring sustainable and profitable carsharing operations.

Future work should explore extensions of this model to encompass the complexity of heterogeneous customers by incorporating socio-demographic characteristics. Moreover, scaling up the study to a more realistic and broader context would enhance the generalizability of the proposed approach. Potential enhancements to this framework can be implemented incrementally, such as expanding the utility model to include more features, incorporating additional alternatives into the choice model, and exploring different pricing mechanisms. These extensions can expand the applicability of this approach to more intricate scenarios. Validating the model's performance through real-world applications and integrating it into a broader multi-modal transportation system are avenues worth exploring to further enhance the practical applicability and impact of the developed framework. Overall, extending this approach to a real-life setting requires real-time price experimentation and monitoring. This can only be achieved by a long-term collaboration between academia and carsharing practitioners. Currently, opportunities for such a collaboration are limited. Nevertheless, we hope that this will be possible in the following years as the field matures, similar to what we observed in airline and retail industries over the last decades.

CRediT authorship contribution statement

Beatriz Brito Oliveira: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Funding acquisition, Formal analysis, Data curation. **Selin Damla Ahipasaoglu:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partially financed by the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project "2022.03138.PTDC".

Appendix A. Notation table

Indices:	
$t, t' = \{0, \dots, T\}$	Indices for the set $\mathcal{T}$ of time periods (e.g. hourly slots per day), where $t = 0$ represents the initial conditions of the time horizon (e.g. week) and "overlaps" with $t = T$ for the previous horizon
$i, j, a = \{1, \dots, A\}$	Indices for the set of areas
$p = \{0, 1, \dots, P\}$	Index for the set $\mathcal{P}$ of promotion levels allowed, where $p = 0$ is the index associated with no promotions being offered
$s = \{1, \dots, S\}$	Index for the set $\mathcal{S}$ of steps for the base prices
Parameters:	
$B(t)$	Time category for time period t (e.g., peak, mid-day, off-peak)
$G(i)$	Area category for area a (e.g., downtown, mobility hub, suburbs)
N	Number of categories for time periods
M	Number of categories for areas
α^0	Attribute-independent parameter of the utility function
$\beta^1_{B(t)}$	Parameter of the utility function related with time category $BT(t)$ of a car's pick-up time t (independent of the price)
$\beta^2_{G(i)}$	Parameter of the utility function related with area category $G(i)$ of a car's pick-up area i (independent of the price)
$\beta^3_{G(j)}$	Parameter of the utility function related with area category $G(j)$ of a car's drop-off area j (independent of the price)
β^4	Parameter of the utility function related with price
K_{ijt}	Time to drive from area i to area j starting at time period t
C_{ij}	Relocation cost per car transferred from area i to area j
W_i	Initial number of cars idle at area i, at the beginning of the time horizon ($t = 0$)
R_{it}	Number of cars on on-going relocations (decided on the previous planning horizon), being transferred to area i, arriving at time t
O_{it}	Number of cars on on-going user trips (decided on the previous planning horizon), being dropped-off in area i at time t
L_s	Length (in time periods) of pricing step s.
Q_{ijt}	Last pricing step for a trip starting in area i at time period t and ending at j. This step can be directly derived from the total time of the trip (DT_{ijt})
B_{its}	Base price (in monetary units) charged per time period, for trips starting in time period t and area i in pricing step s
F_{ijt}	Full price (in monetary units) charged, for trips starting in time period t and area i and ending in area j
V^p_{it}	Promotion value (percent of the full price) associated with promotion level p for time period t and area i, where $V^0_{it} = 0$
D_{ijt}	Total potential demand for trip starting in time period t and area i and ending in area j
Z^p_{ijt}	Parameter pre-calculated for each promotion level and trip to allow linearizing the model. Represents the probability of a user choosing carsharing for each promotion level p offered for trips starting at time period t in area i going to area j

Appendix B. Jenks natural breaks and K-means clustering for the categorizing approach

See Tables B.1 and B.2.

Table B.1
Area categories resulting from K-means Clustering.

Area (a)	Pick-up parameter (β'^2)	Drop-off parameter (β'^3)	Category
0	4.34	2.61	1
1	4.90	2.68	1
2	6.27	2.84	0
3	6.59	2.92	0

Table B.2
Time categories resulting from Jenks Natural Breaks for each categorizing strategy type.

Period (t)	Time parameter (β'^1)	Associated category for each categor. strategy				
		B1	B2	B3	B4	B5
0	1.06	0	0	0	1	4
1	1.16	0	0	1	2	7
2	2.19	1	3	6	12	30
3	2.34	1	3	7	13	32
4	1.49	0	1	3	6	19
5	1.00	0	0	0	0	0
6	1.52	0	1	3	6	20
7	2.92	1	4	9	19	40
8	2.35	1	3	7	13	32
9	1.27	0	1	1	3	12
10	1.05	0	0	0	1	3
11	1.85	0	2	5	9	25
12	1.09	0	0	0	1	6
13	1.40	0	1	2	5	16
14	2.61	1	3	8	16	36
15	2.39	1	3	7	14	33
16	1.82	0	2	5	9	25
17	1.00	0	0	0	0	0
18	1.43	0	1	2	5	17
19	2.29	1	3	7	13	31
20	2.75	1	4	8	17	37
21	1.45	0	1	3	5	18
22	1.05	0	0	0	1	3
23	1.55	0	1	3	6	21
24	1.06	0	0	0	1	4
25	1.37	0	1	2	4	15
26	2.14	1	3	6	12	29
27	2.83	1	4	9	18	38
28	1.49	0	1	3	6	19
29	1.00	0	0	0	0	0
30	1.59	0	1	3	7	21
31	2.94	1	4	9	19	40
32	2.43	1	3	7	15	33
33	1.40	0	1	2	5	16
34	1.05	0	0	0	1	3
35	1.40	0	1	2	5	16
36	1.09	0	0	0	1	6
37	1.31	0	1	2	3	13
38	1.84	0	2	5	9	25
39	2.85	1	4	9	18	39
40	1.31	0	1	2	3	13
41	1.00	0	0	0	0	0
42	1.78	0	2	4	9	24
43	2.92	1	4	9	19	40
44	2.17	1	3	6	12	30
45	1.19	0	0	1	2	9
46	1.04	0	0	0	1	2
47	1.66	0	2	4	7	22
48	1.07	0	0	0	1	5
49	1.22	0	0	1	2	11
50	1.73	0	2	4	8	23
51	3.00	1	4	9	20	41
52	1.65	0	2	4	7	22
53	1.00	0	0	0	0	0

(continued on next page)

Table B.2 (*continued*).

Period (t)	Time parameter (β'^1)	Associated category for each categor. strategy				
		B1	B2	B3	B4	B5
54	1.37	0	1	2	4	15
55	1.97	1	2	5	10	26
56	1.95	1	2	5	10	26
57	1.20	0	0	1	2	10
58	1.05	0	0	0	1	3
59	1.77	0	2	4	8	24
60	1.05	0	0	0	1	3
61	1.18	0	0	1	2	8
62	2.38	1	3	7	14	33
63	2.47	1	3	7	15	34
64	1.83	0	2	5	9	25
65	1.00	0	0	0	0	0
66	1.53	0	1	3	6	20
67	2.06	1	2	6	11	28
68	2.55	1	3	8	16	35
69	1.45	0	1	3	5	18
70	1.03	0	0	0	1	1
71	1.44	0	1	2	5	18
72	1.06	0	0	0	1	4
73	1.16	0	0	1	2	7
74	2.03	1	2	6	11	27
75	3.00	1	4	9	20	41
76	1.74	0	2	4	8	23
77	1.00	0	0	0	0	0
78	1.32	0	1	2	4	13
79	2.08	1	2	6	11	28
80	1.81	0	2	5	9	25
81	1.35	0	1	2	4	14
82	1.04	0	0	0	1	2
83	1.65	0	2	4	7	22

Appendix C. Linear demand model

The main feature of the linear demand (LD-MILP) approach is that it does not update prior estimates of demand and assumes that demand varies linearly with price. This approach is based on the mathematical optimization model (7)–(15), but differs in the way the demand parameter (D_{ijt}) is defined. The linear demand model used in these experiments is based on a stepwise knowledge of the demand for the highest and lowest promotion levels. It assumes the same values for all trip types, depending on the promotion level charged, and is calculated for regular and extended promotions. We consider the highest and lowest demand levels to correspond to the upper and lower promotion levels, while the middle promotion level is assigned the average demand under perfect information. For the remaining levels, the demand values result from a linear interpolation between the lower and middle levels or the middle and upper levels, depending on their bounds. Fig. C.1 shows the values used for regular and extended promotions.

Appendix D. Full results

See Tables D.1 and D.2.

Appendix E. Milan dataset results for extended promotions

See Table E.1

Data availability

The data is shared in a link in the paper.

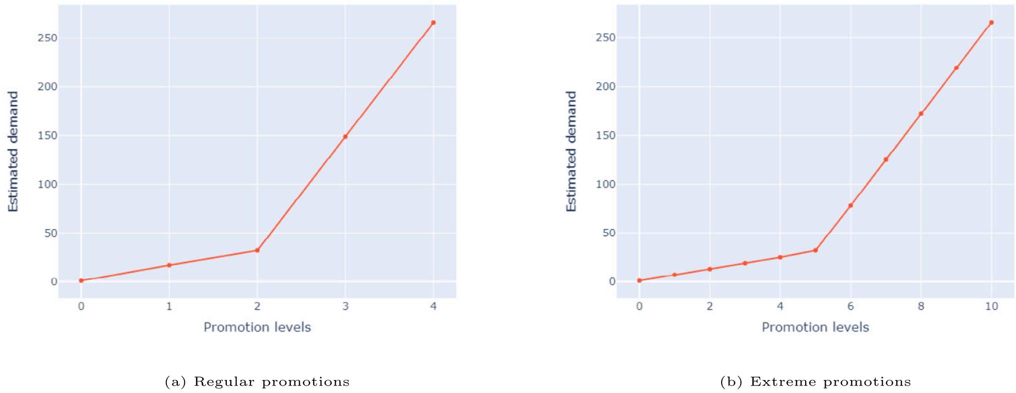


Fig. C.1. Linear price-demand model parametrization.

Table D.1
Results for the MILP model and realized profit per iteration and per approach, considering regular promotions.

Iteration	Proposed approach			
	Estimated profit	Time to run (sec)	MILP gap	Realized profit
1	62,139	600	0.05%	58,766
2	62,050	600	0.07%	58,893
3	62,020	600	0.06%	58,566
4	62,042	600	0.06%	58,655
5	62,036	600	0.07%	58,633
average	**62,057**	**600**	**0.06%**	**58,703**
Iteration	**Linear demand approach**			
	Estimated profit	**Time to run (sec)**	**MILP gap**	**Realized profit**
1	64,743	600	0.35%	57,421
2	64,547	600	0.36%	57,524
3	64,577	600	0.29%	57,270
4	64,633	600	0.27%	57,493
5	64,625	600	0.32%	57,427
average	**64,625**	**600**	**0.32%**	**57,427**
Iteration		**Perfect Information**		
		Time to run (sec)	**MILP gap**	**Realized profit**
1		300	0.00%	76,607
2		315	0.00%	76,677
3		315	0.00%	76,558
4		316	0.00%	76,582
5		316	0.00%	76,566
average		**312**	**0.00%**	**76,598**

Table D.2
Results for the MILP model and realized profit per iteration and per approach, considering extended promotions.

Iteration	Proposed approach			
	Estimated profit	Time to run (sec)	MILP gap	Realized profit
1	47,768	600	0.09%	47,520
2	47,745	600	0.09%	47,713
3	47,741	600	0.10%	47,523
4	47,758	600	0.09%	47,466
5	47,757	600	0.09%	47,392
average	**47,754**	**600**	**0.09%**	**47,523**
Iteration	**Linear demand approach**			
	Estimated profit	**Time to run (sec)**	**MILP gap**	**Realized profit**
1	43,219	600	1.11%	40,079
2	43,066	600	0.97%	40,328
3	43,075	600	1.10%	39,867
4	43,055	600	1.14%	40,084
5	43,104	600	1.08%	40,089
average	**43,104**	**600**	**1.08%**	**40,089**
Iteration		**Perfect Information**		
		Time to run (sec)	**MILP gap**	**Realized profit**
1		19	0.00%	66,287
2		19	0.00%	66,297
3		20	0.00%	66,062
4		21	0.00%	66,279
5		21	0.00%	66,036
average		**20**	**0.00%**	**66,192**

Table E.1
Decisions on pricing (distribution of trip types per promotion level selected) and relocations, and resulting service level metrics for extended promotions (average for all iterations) for the real dataset.

Promotion level	BL-MILP	LD-MILP
0	36%	43%
1	0%	0%
2	1%	28%
3	12%	0%
4	5%	2%
5	42%	0%
6	4%	10%
7	0%	16%
8	0%	0%
9	0%	0%
10	0%	0%
Demand	155	174
Trips served	153	171
Service level	98%	99%
Relocations	21	8

References

Archetti, C., Bruglieri, M., Guastaroba, G., Speranza, M.G., 2023. The benefit of complete trip information in free-floating carsharing systems. Eur. J. Transp. Logist. 12 (July 2022), http://dx.doi.org/10.1016/j.ejtl.2023.100110.

Balac, M., Becker, H., Ciari, F., Axhausen, K.W., 2019. Modeling competing free-floating carsharing operators – A case study for Zurich, Switzerland. Transp. Res. C 98 (May 2018), 101–117. http://dx.doi.org/10.1016/j.trc.2018.11.011.

Banerjee, S., Freund, D., Lykouris, T., 2021. Pricing and optimization in shared vehicle systems: An approximation framework. Oper. Res. 3, http://dx.doi.org/10.1287/opre.2021.2165.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, R. (Ed.), Handbook of Transportation Science. International Series in Operations Research & Management Science, Vol 23.. Springer, MA, pp. 5–33. http://dx.doi.org/10.1007/978-1-4615-5203-1_2.

Benjaafar, S., Gao, X., Shen, X., Zhang, H., 2023. Online learning for pricing in on-demand vehicle sharing networks. SSRN 4344364, http://dx.doi.org/10.2139/ssrn.4344364.

Bierlaire, M., 2023. A short introduction to Biogeme. Technical report, (TRANSP-OR 230620), Transport Mobility Laboratory, ENAC, EPFL, https://transp-or.epfl.ch/documents/technicalReports/Bier23.pdf.

Boyacı, B., Zografos, K.G., Geroliminis, N., 2017. An integrated optimization-simulation framework for vehicle and personnel relocations of electric carsharing systems with reservations. Transp. Res. B 95, 214–237. http://dx.doi.org/10.1016/j.trb.2016.10.007.

Braverman, A., Dai, J.G., Liu, X., Ying, L., 2019. Empty-car Routing in Ridesharing systems. Oper. Res. 67, 1437–1452. http://dx.doi.org/10.1287/opre.2018.1822.

Chen, Q., Lei, Y., Jasin, S., 2024. Real-time spatial–intertemporal pricing and relocation in a ride-hailing network: near-optimal policies and the value of dynamic pricing. Oper. Res. 72, 2097–2118. http://dx.doi.org/10.1287/opre.2022.2425.

Golalikhani, M., Oliveira, B.B., Carravilla, M.A., Oliveira, J.F., Antunes, A.P., 2021. Carsharing: A review of academic literature and business practices toward an integrated decision-support framework. Transp. Res. E 149 (February), 102280. http://dx.doi.org/10.1016/j.tre.2021.102280.

Hosseini, M., Milner, J., Romero, G., 2024. Dynamic relocations in car-sharing networks. Oper. Res. 72, http://dx.doi.org/10.1287/opre.2021.0062.

Huang, K., An, K., Correia, G.H.d., Rich, J., Ma, W., 2021. An innovative approach to solve the carsharing demand-supply imbalance problem under demand uncertainty. Transp. Res. C 132 (July), 103369. http://dx.doi.org/10.1016/j.trc.2021.103369.

Huang, K., An, K., Rich, J., Ma, W., 2020. Vehicle relocation in one-way station-based electric carsharing systems: A comparative study of operator-based and user-based methods. Transp. Res. E 142 (July), 102081. http://dx.doi.org/10.1016/j.tre.2020.102081.

Huang, K., Correia, G.H.d., An, K., 2018. Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand. Transp. Res. C 90 (February), 1–17. http://dx.doi.org/10.1016/j.trc.2018.02.020.

Illgen, S., Höck, M., 2019. Literature review of the vehicle relocation problem in one-way car sharing networks. Transp. Res. B 120, 193–204. http://dx.doi.org/10.1016/j.trb.2018.12.006.

Jorge, D., Correia, G.H., Barnhart, C., 2014. Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems. IEEE Trans. Intell. Transp. Syst. 15 (4), 1667–1675. http://dx.doi.org/10.1109/TITS.2014.2304358.

Jorge, D., Molnar, G., de Almeida Correia, G.H., 2015. Trip pricing of one-way station-based carsharing networks with zone and time of day price variations. Transp. Res. B 81, 461–482. http://dx.doi.org/10.1016/j.trb.2015.06.003.

Khamis, N., Sin, T.C., Hock, G.C., 2018. Segmentation of residential customer load profile in peninsular Malaysia using jenks natural breaks. In: 2018 IEEE 7th International Conference Power Energy, PECon 2018. IEEE, pp. 128–131. http://dx.doi.org/10.1109/PECON.2018.8684113.

Liu, Y., Xie, J., Chen, N., 2022. Stochastic one-way carsharing systems with dynamic relocation incentives through preference learning. Transp. Res. E 166 (August), 102884. http://dx.doi.org/10.1016/j.tre.2022.102884.

Lu, R., Correia, G.H.d., Zhao, X., Liang, X., Lv, Y., 2021. Performance of one-way carsharing systems under combined strategy of pricing and relocations. Transp. B 9 (1), 134–152. http://dx.doi.org/10.1080/21680566.2020.1819912.

McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, pp. 198–271.

Oliveira, B.B., Carravilla, M.A., Oliveira, J.F., 2018. Integrating pricing and capacity decisions in car rental: A matheuristic approach. Oper. Res. Perspect. 5, 334–356. http://dx.doi.org/10.1016/j.orp.2018.10.002, URL https://linkinghub.elsevier.com/retrieve/pii/S2214716018302707.

Ortúzar, J.d.D., Willumsen, L., 2011. Modelling Transport, 4th ed. John Wiley and Sons, Ltd., Chichester, UK.

Pantuso, G., 2022. Exact solutions to a carsharing pricing and relocation problem under uncertainty. Comput. Oper. Res. 144 (March), 105802. http://dx.doi.org/10.1016/j.cor.2022.105802.

Ren, S., Luo, F., Lin, L., Hsu, S.C., LI, X.I., 2019. A novel dynamic pricing scheme for a large-scale electric vehicle sharing network considering vehicle relocation and vehicle-grid-integration. Int. J. Prod. Econ. 218 (July), 339–351. http://dx.doi.org/10.1016/j.ijpe.2019.06.020.

Samie, S., Rezaee, B., 2022. Dynamic discrimination pricing and freelance drivers to rebalance mixed-fleet carsharing systems. IEEE Trans. Intell. Transp. Syst. 23 (12), 24738–24752. http://dx.doi.org/10.1109/TITS.2022.3206413.

Shaheen, S.A., Chan, N.D., Micheaux, H., 2015. One-way carsharing's evolution and operator perspectives from the americas. Transportation 42 (3), 519–536. http://dx.doi.org/10.1007/s11116-015-9607-0.

Soppert, M., Steinhardt, C., Müller, C., Gönsch, J., Bhogale, P.M., 2023. Matching functions for free-floating shared mobility system optimization to capture maximum walking distances. European J. Oper. Res. 305 (3), 1194–1214. http://dx.doi.org/10.1016/j.ejor.2022.06.058.

Train, K.E., 2003. Discrete choice methods with simulation. In: Discrete Choice Methods with Simulation, vol. 9780521816. Cambdrige University Press, pp. 1–334. http://dx.doi.org/10.1017/CBO9780511753930.

Wang, L., Ma, W., 2019. Pricing approach to balance demands for one-way car-sharing systems. In: 2019 IEEE Intelligent Transportation System Conference ITSC 2019. IEEE, pp. 1697–1702. http://dx.doi.org/10.1109/ITSC.2019.8917414.

Yang, S., Wu, J., Sun, H., Qu, Y., Wang, D.Z., 2022. Integrated optimization of pricing and relocation in the competitive carsharing market: A multi-leader-follower game model. Transp. Res. C 138 (June 2021), 103613. http://dx.doi.org/10.1016/j.trc.2022.103613.