This is the Pre-Published Version.

This is an Accepted Manuscript of an article published by Taylor & Francis in IISE transactions on 22 Oct 2024 (published online), available at: https://doi.org/10.1080/24725854.2024.2404555.

Integrated Vehicle Allocation and Relocation for Shared Micromobility under Competition and Demand Uncertainty

Ziliang Jin [a], Kai Pan [b,*], Zuo-Jun Max Shen [c], and Wenxin Xu [d]

[a] School of Economics and Management, Southeast University, Nanjing, China, ziliang.jin@seu.edu.cn

[b] Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong, kai.pan@polyu.edu.hk

[c] Faculty of Engineering and Faculty of Business and Economics, The University of Hong Kong, Hong Kong, maxshen@berkeley.edu

[d] Darla Moore School of Business, University of South Carolina, Columbia, SC 29208, USA, wenxin.xu@moore.sc.edu

* Corresponding author

Abstract

This paper aims to understand the shared micromobility firms' operations under competition and provide managerial guidance to the firms and the regulator. We consider two shared micromobility firms competing in the same service area, each providing micromobility vehicles to satisfy uncertain demands. Concerning allocation restrictions by the regulator, we propose an innovative capacity-sharing agreement between the two firms. Each firm solves an integrated vehicle allocation and relocation problem, modeled as a two-stage stochastic program on a spatial-temporal network. We explore the optimality condition of each firm's decision-making and seek a Nash equilibrium by optimizing certain objectives over the joint optimality conditions of both firms. We prove that capacity sharing helps reduce the total demand loss in the system. We perform extensive numerical experiments based on real data to obtain managerial insights. We find that regulator restrictions impact firms' profitability and service level. After introducing capacity sharing, one firm may act like a free rider that relies on the vehicles transferred from her opponent. Meanwhile, many vehicles are shared in periods and regions with high trip demands. Capacity sharing can reduce the number of relocated vehicles by serving as a substitution for relocation and also improves the firms' profitability.

Keywords: Shared Micromobility; Allocation and Relocation; Two-stage Stochastic Programming; Competition; Capacity Sharing.

1 Introduction

A *shared micromobility system* consists of lightweight vehicles (e.g., bikes, e-bikes, e-scooters) and well supports urban short-distance travel (e.g., last-mile transportation), alleviating city congestion because of its eco-friendliness and sustainability (McKinsey, 2021). Note that 60% of today's total passenger miles traveled in China, Europe, and the U.S. are less than 8 kilometers (McKinsey, 2019), and the energy prices and public interest in social distancing are growing. Thus, the shared micromobility system becomes a significant transportation solution for society.

The shared micromobility system allows consumers to conveniently pick up and drop off vehicles in any service region at any time, offering short-term self-service rental programs (Fishman et al., 2013). A consumer can rent a micromobility vehicle by paying a low rental cost (much lower than the ownership and maintenance costs) for a short-distance trip. Motivated by such convenience, over 1,000 cities worldwide have established similar programs (Wang and Lindsey, 2019). For example, nearly seventy bike-sharing companies were operating in China as of July 2017 (Sago, 2020); six e-scooter companies (i.e., Lime, Bird, Tier Mobility, Wind Mobility, Flash, and Hive) are competing in Vienna, and five companies (i.e., Lime, Bird, Lyft, Skip by Helbiz, and Spin) in Washington, DC (Schellong et al., 2019).

A shared micromobility system differs from other shared mobility systems (e.g., ride sourcing) mainly because of the following three features. First, shared micromobility services are not provided on-demand, like ride sourcing, due to the lack of self-regulated owners for each vehicle. Each micromobility vehicle should stay ready for the coming consumers, significantly challenging the operators to allocate and relocate vehicles efficiently for high service quality and profitability. Second, the convenience of free pick-up and drop-off often creates vehicle supply and demand imbalance in different regions (Jin et al., 2023). For instance, consumers often drop off micromobility vehicles near subway stations, leading to an oversupply of vehicles, while there may be insufficient vehicles in regions far away from central business regions. This can significantly mitigate the firm's operational efficiency, service quality, and profitability. Thus, a shared micromobility firm (referred to as "she") must relocate her vehicles across the service regions efficiently. Third, the shared micromobility firm often owns the vehicles and bears high investment and operational costs with low rental prices (Hasija et al., 2020), and is further responsible for carefully operating the system to match supply with demand. The firm cannot simply raise rental prices to cover these high costs because consumers can easily switch to other mobility services. Thus, efficient vehicle allocation and relocation across regions to ensure profitability and service quality are significantly important for a shared micromobility firm.

Currently, multiple shared micromobility firms operate in a city. The fierce competition among firms brings additional challenges to both the firms and the regulator. On the one hand, such a competition, along with the aforementioned three features of the shared micromobility system, requires the firms to efficiently perform initial allocation and subsequent relocation of vehicles to ensure commercial sustainability. On the other hand, the regulator faces the challenge of regulating the shared micromobility industry to avoid too many vehicles blocking traffic and causing safety issues while maintaining the benefits of the shared micromobility services to city residents. In many cities worldwide, e.g., Barcelona (Valdivia, 2020), San Diego (Hargrove and Ahn, 2022), Chicago (City of Chicago, 2022), and Boston (American Legal Publishing, 2019), regulators limit the number of shared micromobility vehicles allocated in the city via licenses issued to the firms. Each firm should apply for a portion of these licenses. In addition, the City of Chicago requires that the licensed vehicles be allocated relatively fairly in the service regions based on consumer demands to ensure that vehicles can be available to all city residents. They further periodically adjust the number of licensed vehicles by reviewing the utilization rate (City of Chicago, 2022). Furthermore, they divide the service area into different types of areas, each with sub-areas. The regulator sets a specific allocation limit for each sub-area.

With a limited number of vehicles due to the regulator's restriction on vehicle allocation, shared micromobility firms are challenged to provide consumers with high-quality services. To address this challenge, we suggest an innovative *capacity-sharing* agreement between the firms, under which a firm can share spare capacity (i.e., micromobility vehicles) for a fee with her opponent when both are competing in the same market. To the best of our knowledge, we are among the first to integrate the initial allocation and subsequent relocation of shared micromobility vehicles under firm competition, regulator restrictions, and demand uncertainty. For such a case, we are interested in the following research questions: How should a shared micromobility operator integrate vehicle allocation and relocation? How do the firm competition and regulator restrictions affect the firms' performance? How does capacity sharing affect the firm's profitability and operations? Should the regulator intervene in the shared micromobility industry to balance social welfare and the industry's economic viability, and if so, what are possible approaches?

To answer these questions, we consider two shared micromobility firms competing in the same service area that is divided into multiple regions (He et al., 2017; Qi and Shen, 2019). Each firm provides a fleet of micromobility vehicles to satisfy consumer demands over an operational horizon with multiple periods. Motivated by industrial practices, we consider uncertain consumer demands, and the market consists of two types of consumers: *loyal consumers* and *disloyal consumers*. A loyal consumer of a firm always chooses that firm when renting micromobility vehicles, whereas a disloyal consumer may choose either firm. The reasons for a loyal consumer's preference vary. For instance, he may prefer the firm's vehicles if the two firms offer different ones, or he may be ineligible to rent the other firm's vehicles if he has not paid a deposit fee yet.

We formulate a two-stage model. In the first stage, each firm builds her capacity by initially allocating micromobility vehicles across service regions without knowing the actual demands in each period. The regulator limits the number of vehicles allocated in each service region by the two firms together. We call this stage *the capacity-development stage*. In the second stage, after the demands are realized, each firm operates her vehicles by subsequently relocating vehicles across service regions in each period to rebalance the vehicles toward satisfying the demands. We call this stage *the operation stage*. Moreover, following the study by Kabra et al. (2020) that examines the impact of bike accessibility and availability upon bike-share ridership, we consider that the consumer demands partially depend on the vehicle capacity (i.e., the number of allocated vehicles). The mismatch between vehicle supply and demand may cause parking costs for idle vehicles and lost-sale penalties. Each firm maximizes her total expected profit over the entire operational horizon, including the revenue from serving consumers and the costs incurred in the two stages. We make the following contributions:

(i) Each shared micromobility firm solves an integrated vehicle allocation and relocation problem and provides a Nash equilibrium. Each firm's decision-making problem is formulated as a two-stage stochastic program on a spatial-temporal network (Mahmoudi and Zhou, 2016; Lu et al., 2017). In the capacity-development stage (1st stage), the firm decides on initial vehicle allocation for service regions. In the operation stage (2nd stage), the firms make subsequent vehicle relocation as recourse across service regions to match supply with realized demand in each period. We also consider transferred micromobility vehicles when both firms share their capacities (i.e., vehicles). To the best of our knowledge, our formulated spatial-temporal network is among the first to incorporate the flows between the two firms.

(ii) We explore the optimality condition of each firm's decision-making and seek the Nash equilibrium by optimizing certain objectives (i.e., criteria for selecting an equilibrium) over the joint optimality conditions of both firms. Specifically, we consider two criteria preferred by the regulator and the firms, respectively: minimizing the total demand loss and minimizing the total number of allocated vehicles.

(iii) We propose an innovative *capacity-sharing* agreement between the shared micromobility firms. Under this agreement, each firm can share abundant vehicles with her opponent for a fee. We demonstrate the feasibility and effectiveness of the capacity-sharing agreement.

(iv) We perform numerical experiments based on real data collected from Citi Bike (2022) and obtain the following managerial insights for the regulator and the firms. The regulator restrictions impact firms' profitability and service level. The competition benefits the firm with fewer loyal consumers by increasing her profit. After introducing capacity sharing, one firm may act like a free rider that allocates fewer vehicles and asks for vehicles transferred from her opponent if needed. Meanwhile, many vehicles are shared in periods and regions with high trip demands. Capacity sharing can reduce the number of relocated vehicles by serving as a substitution for relocation and also improves the firms' profitability.

The remainder of the paper is organized as follows. Section 2 reviews related studies. Section 3 introduces the problem formulation. We conduct numerical experiments in Section 4 based on real data from Citi Bike (2022). Section 5 concludes the paper.

2 Literature Review

Our work builds on the prior Operations Management (OM) studies on shared mobility. With the advanced development of disruptive technologies and data analytics, the smart city paradigm is shifting from new technology adoption to decision optimization, focusing on improving operational efficiency (Qi and Shen, 2019; Mak, 2022). In line with this shift, a growing body of OM literature studies smart city problems.

Shared mobility, a component of smart cities, proliferates in industry and has also received attention in recent literature. Many studies on shared mobility focus on ride sourcing, with which people rent cars from individual car owners (often also drivers) for trips (e.g., see Qi et al. 2018, Yu and Shen 2020, Benjaafar et al. 2022, and Moug et al. 2023). Shared micromobility, complementary to ride sourcing, also brings challenges and opportunities to OM researchers (Hasija et al., 2020). Shared micromobility is fundamentally different from ride sourcing in the ownership of assets. Shared micromobility firms are asset-based, owning micromobility vehicles (e.g., bikes or scooters), whereas ride-sourcing firms are often two-sided platforms that provide services to match supply with demand (Chen et al., 2023). Recently, Kabra et al. (2020) and He et al. (2021a) study how the bike supply impacts end-user behavior. Our work is more related to Shu et al. (2013), and they formulate two linear programming models to optimize the number of bikes in each dock and the fleet operation, respectively. However, the initial allocation and fleet operations are not independent, especially for the new entrants of the shared micromobility market. In this paper, each shared micromobility firm integrates the capacity development (i.e., initial vehicle allocation) and fleet operations (i.e., subsequent vehicle relocation) in a two-stage stochastic program. Unlike another relevant paper by Jin et al. (2023), we consider such an integration under firm competition and consumer demands depend on capacity development decisions.

Our paper extends the literature on fleet operation in shared mobility. There is a strand of papers that analyze fleet operation in ride sourcing (see, e.g., Yu and Shen 2020 and Moug et al. 2023). The fleet operation in shared micromobility, however, depends on capacity decisions (e.g., the number of allocated bikes) due to the difference in asset ownership. As mentioned in Section 1, shared micromobility firms must carefully determine capacity development and operation together to achieve economic viability and sustainability. It is also feasible to consider capacity decisions and study sharing capacities in shared micromobility because the firms can easily decide and adjust the capacity as they often own the assets. In contrast, ride-sourcing firms have to use incentive instruments to affect the capacity, e.g., motivating more drivers to provide service by increasing their salaries. Several works dedicate to the fleet operation in shared micromobility (e.g., see Nair and Miller-Hooks 2011, Zhang et al. 2017, and Schuijbroek et al. 2017). For instance, He et al. (2021b) formulate a model wherein a two-sided matching platform (e.g., a free-float bike-sharing company) determines the spatial allocation of parking spaces and the design of incentive instruments that can affect the supply (e.g., price and rewards to users). However, to the best of our knowledge, none of these studies explicitly considers a firm's capacity decision (e.g., bike allocation), which is more critical in shared micromobility than in other shared mobility (e.g., ride sourcing). Motivated by the necessity and feasibility, our work contributes to this stream of literature by integrating the capacity decision and fleet operation for shared micromobility.

Our study contributes to the steam of literature on competition in shared mobility. The prior works mainly focus on the competition in ride sharing (e.g., see Nikzad 2017 and Zhang et al. 2022). The competition in shared micromobility is vastly observed in practice while receiving limited attention in the literature. Most related studies on competition in shared micromobility focus on the demand side. Cao et al. (2021) use staggered entry of two deckles bike-sharing firms and observe a positive network effect in demand between the firms. Reck et al. (2022) analyze the impact of competition on demand between different mode choices (i.e., dockless e-scooters and e-bikes and docked e-bikes and bikes). Our work contributes to this stream of literature by considering the supply side—the capacity decisions. While there is extensive literature on capacity investment under competition and uncertainty (see, e.g., Van Mieghem 2003, Chevalier-Roignant et al. 2011, and Jiang et al. 2020), to the best of our knowledge, we are among the first to consider competitive capacity investment and operational decisions of shared micromobility firms under demand uncertainty while also considering capacity sharing between firms.

3 Model

We consider two competing shared micromobility firms (indexed as $k \in \mathcal{K} = \{A, B\}$) that just expand their business into a city. Each firm provides shared micromobility service for a set of regions $\mathcal{M} = \{1, 2, \ldots, M\}$ in each period $t \in \mathcal{T} = \{0, 1, \ldots, T\}$.

3.1 Sequence of Events

We build a two-stage model, where the first stage is the *capacity-development stage* and the second stage is the *operation stage*. We introduce the sequence of events and the firms' decisions below.

(a) Capacity-development stage: Each firm $k \in \mathcal{K}$ simultaneously decides the number of micromobility vehicles to be allocated in each region $i \in \mathcal{M}$, denoted by x_i^k. We assume that each firm k is endowed with a budget that can afford at most $\hat{x}^k$ vehicles. In addition, based on approximated trip demands in each region, the regulator (e.g., the city government) restricts the numbers of micromobility vehicle licensing plates and available parking spaces to avoid traffic chaos. Thus, both firms together allocate at most $\bar{x}_i$ vehicles in each region $i \in \mathcal{M}$.

(b) Operation stage: In each period $t \in \mathcal{T}$, consumers arrive to rent micromobility vehicles, and each firm simultaneously decides (i) the number of vehicles that fulfill the consumers' rental requests; (ii) the number of vehicles that are relocated across regions to balance the vehicle inventories with the demands; and (iii) the number of vehicles that stay idle in each region.

3.2 Vehicle Movement: A Spatial-temporal Network

We model the vehicle movements among regions and across periods at the operation stage as *flows* in a spatial-temporal network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ is the set of directed arcs on the network (see Figure 1). A node $n_{i,t}^k \in \mathcal{N}$ represents service region $i \in \mathcal{M}$ in period $t \in \mathcal{T}$ for firm $k \in \mathcal{K}$. For each firm k, the *flow* on the directed arc $(n_{i,t}^k, n_{i',t'}^k) \in \mathcal{E}$ represents the number of micromobility vehicles moving from region i in period t (i.e., node $n_{i,t}^k$) to region i' in period t' (i.e., node $n_{i',t'}^k$).

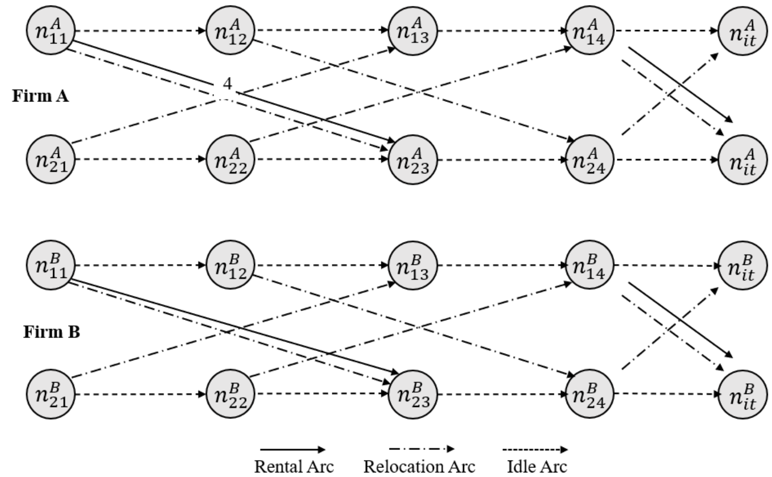


Figure 1. Spatial-Temporal Network $\mathcal{G}$

Based on how vehicles are moved between two nodes, we define three types of arcs in $\mathcal{E}$ for each firm $k \in \mathcal{K}$: (i) *Rental arcs*: The flow on each rental arc $(n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}}$ represents the number of rental trips from region i in period t to region $i' \neq i$ in period $t' = t + \ell_{i,i'}$, where $\ell_{i,i'}$ represents the number of periods that a micromobility vehicle rider needs to move from region i to region i' and $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}} = \{\mathcal{M} \times \mathcal{M} \times \mathcal{T} \times \mathcal{T} \mid i \neq i',\ t' = t + \ell_{i,i'}\}$. (ii) *Idle arcs*: The flow of each idle arc $(n^k_{i,t}, n^k_{i,t+1}) \in \mathcal{E}^k_{\mathrm{I}}$ represents the number of idling vehicles in region i from period t to period $t+1$, where $(i,t) \in \mathcal{Z}_{\mathrm{I}} = \{\mathcal{M} \times \mathcal{T} \mid t \in \mathcal{T} \setminus \{T\}\}$. (iii) *Relocation arcs*: The flow of each relocation arc $(n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{L}}$ represents the number of vehicles relocated by the firm from region i in period t to region $i' \neq i$ in period $t' = t + r_{i,i'}$, where $r_{i,i'}$ represents the number of periods that a vehicle needs to be relocated from region i to region i' and $(i, i', t, t') \in \mathcal{Z}_{\mathrm{L}} = \{\mathcal{M} \times \mathcal{M} \times \mathcal{T} \times \mathcal{T} \mid i \neq i',\ t' = t + r_{i,i'}\}$. Let $\mathcal{A} = \{\mathrm{R}, \mathrm{I}, \mathrm{L}\}$. We have $\mathcal{E}^k = \cup_{\forall a \in \mathcal{A}} \mathcal{E}^k_a$, $\mathcal{E}^k_{a_1} \cap \mathcal{E}^k_{a_2} = \emptyset$ for given $k \in \mathcal{K}$, $a_1 \neq a_2$, $a_1, a_2 \in \mathcal{A}$, and $\mathcal{E}^A_a \cap \mathcal{E}^B_a = \emptyset$ for given $a \in \mathcal{A}$. Note that both Firm A and Firm B use similar micromobility vehicles and relocation methods and share the geographic regions in $\mathcal{M}$, the values of $\ell_{i,i'}$ and $r_{i,i'}$ are firm-independent for any $i, i' \in \mathcal{M}$. Thus, for any arc $(n^A_{i,t}, n^A_{i',t'}) \in \mathcal{E}^A$, we have a corresponding arc $(n^B_{i,t}, n^B_{i',t'}) \in \mathcal{E}^B$, and these two arcs represent two flows from region i in period t to region i' in period t' for Firm A and Firm B, respectively.

3.3 Consumer Types

We consider two types of consumers: *loyal consumers* and *disloyal consumers*. A loyal consumer of a firm always chooses that firm to rent micromobility vehicles from, regardless of the rental prices. The reasons for such a preference are various. For instance, the consumer may prefer the firm's vehicles if the two firms offer different ones, or the consumer is ineligible to rent the other firm's vehicles if he has not paid a deposit fee yet. Let $\alpha_k \in [0,1]$ and $\alpha_{-k} \in [0,1]$ denote the percentage of consumers who are loyal to firm k and firm $-k$, respectively. We assume that the remaining consumers (i.e., a percentage of $1 - \alpha_A - \alpha_B \geq 0$ of all the consumers) are disloyal to both firms, and they choose the firm with more available vehicles. The assumption comes from the phenomenon that many consumers are more likely to choose a firm due to easier access to the vehicles (Kabra et al., 2020) instead of lower rental prices because the price difference is usually insignificant.

3.4 Mathematical Formulation

We formulate the decision-making model of each firm $k \in \mathcal{K}$ as a two-stage stochastic program. In the *capacity-development stage*, each firm $k \in \mathcal{K}$ makes the vehicle allocation decision $\mathbf{x}^k = [x_1^k, x_2^k, \ldots, x_M^k]^\top$ such that the following constraints are satisfied:

$$\sum_{i \in \mathcal{M}} x_i^k \leq \hat{x}^k; \quad x_i^k + x_i^{-k} \leq \bar{x}_i, \, \forall\, i \in \mathcal{M}; \quad x_i^k \geq 0, \, \forall\, i \in \mathcal{M}. \tag{1}$$

Given firm $-k$'s decision $\mathbf{x}^{-k}$, we let $\mathcal{X}^k(\mathbf{x}^{-k}) = \left\{\mathbf{x}^k \in \mathbb{R}_+^M \;\middle|\; (1)\right\}$. Thus, the vehicle allocation decision of one firm depends on that of the other. Note that the allocation decision $\mathbf{x}^k$ is a planning decision that should be made prior to the operation (e.g., one day before offering services), as vehicle preparation (e.g., procurement and shipment) is time-consuming. It is generally unnecessary to decide vehicle allocations for *each period* within the operational horizon in advance because we can adjust the number of vehicles in each region and period through relocation during operations. Meanwhile, making allocation decisions for each period (e.g., 15 minutes) is not cost-effective and practical because such decisions imply that the fleet size is dynamically adjusted every period. This necessitates frequent transactions for the purchase or disposal of vehicles to align with the updated allocations, while the frequent dynamic adjustment of allocations across various periods is generally expensive. Furthermore, it is not practical in general to complete the purchase or disposal of vehicles within each short time window (e.g., 15 minutes).

In the *operation stage*, time-dependent demands arrive at each region $i \in \mathcal{M}$ at the start of each period $t \in \mathcal{T} \setminus \{0, T\}$. For each rental arc $e = (n^k_{i,t}, n^k_{i',t+\ell_{i,i'}}) \in \mathcal{E}^k_{\mathrm{R}}$, let $\overline{w}^k_e$ denote the demand (i.e., the number of consumers) from region i in period t to region i' in period $t + \ell_{i,i'}$. The firm receives a revenue $h^k_{\mathrm{R}} \geq 0$ whenever it completes serving a consumer. If the number of micromobility vehicles available in region i cannot satisfy the demands of the region in period t, the unsatisfied demand for $e = (n^k_{i,t}, n^k_{i',t+\ell_{i,i'}}) \in \mathcal{E}^k_{\mathrm{R}}$ is lost and the firm k bears a penalty cost $h^k_{\mathrm{P}} \geq 0$ per consumer lost. If the number of consumers is fewer than the number of vehicles in region i in period t, an idle cost $h^k_{\mathrm{I}} \geq 0$ (e.g., the maintenance cost of idle vehicles) is incurred for each idling vehicle on the arc $e = (n^k_{i,t}, n^k_{i,t+1}) \in \mathcal{E}^k_{\mathrm{I}}$. When each firm $k \in \mathcal{K}$ relocates a vehicle, it pays a cost $h^k_{\mathrm{L}} \geq 0$, including labor and trucking costs.

Anticipating demand uncertainty in the operation stage, each firm $k \in \mathcal{K}$ makes her initial vehicle allocation decision $\mathbf{x}^k$ and subsequent relocation decisions to maximize her expected profit. The profit equals the total revenue minus the total cost. The total cost includes the initial vehicle allocation cost $\mathbf{c}^\top \mathbf{x}^k$ in the capacity-development stage (where $\mathbf{c}^k = [c^k_1, c^k_2, \ldots, c^k_M]^\top$) and demand-loss penalty, idle, and vehicle relocation costs in the operation stage. Given the initial vehicle allocation $\mathbf{x}^k$, let $\varphi^k(\mathbf{x}^k)$ denote the optimal expected net cost of the operation stage (i.e., the total cost minus the total revenue of the operation stage). Thus, maximizing the firm k's total expected profit is equivalent to minimizing the summation of $\mathbf{c}^\top \mathbf{x}^k$ and $\varphi^k(\mathbf{x}^k)$. Each firm $k \in \mathcal{K}$ optimizes her vehicle allocation decisions by solving the following problem:

$$-\Psi^k = \min_{\mathbf{x}^k} \left\{ {\mathbf{c}^k}^\top \mathbf{x}^k + \varphi^k(\mathbf{x}^k) \;\middle|\; \mathbf{x}^k \in \mathcal{X}^k(\mathbf{x}^{-k}) \right\}, \qquad (\mathcal{D}^k(\mathbf{x}^{-k}))$$

where $\Psi^k \geq 0$ represents the total expected profit of each firm $k \in \mathcal{K}$ and $\varphi^k(\mathbf{x}^k)$ is realized in the operation stage when each firm $k \in \mathcal{K}$ relocates her micromobility vehicles across regions. Problem ($\mathcal{D}^k(\mathbf{x}^{-k})$) is an integrated vehicle allocation and relocation problem. Solving problem ($\mathcal{D}^k(\mathbf{x}^{-k})$) for each firm $k \in \mathcal{K}$ simultaneously gives us a Nash equilibrium. We now describe the details of $\varphi^k(\mathbf{x}^k)$ below.

First, we discuss the rental demand of each firm $k \in \mathcal{K}$. For any given time-space range $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$, (i) we have $e_A = (n^A_{i,t}, n^A_{i',t'}) \in \mathcal{E}^A_{\mathrm{R}}$ and $e_B = (n^B_{i,t}, n^B_{i',t'}) \in \mathcal{E}^B_{\mathrm{R}}$ sharing the same time-space range; (ii) we let $Y^k_{i,i',t,t'}$ denote the number of loyal consumers of firm $k \in \mathcal{K}$ on each rental arc $e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}}$; (iii) we let $Y^0_{i,i',t,t'}$ denote the total number of disloyal consumers on the arcs e_A and e_B together. Thus, the total market demand (i.e., the total number of consumers) for any $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$ (i.e., on the arcs $e_A = (n^A_{i,t}, n^A_{i',t'})$ and $e_B = (n^B_{i,t}, n^B_{i',t'})$ together) is $Y^A_{i,i',t,t'} + Y^B_{i,i',t,t'} + Y^0_{i,i',t,t'}$. Recall that $\sum_{i\in\mathcal{M}} \bar{x}_i$ is the maximum number of micromobility vehicles allowed in the service area by the regulator's restriction, and it approximates the size of trip demands in the area (City of Chicago, 2022). We thus assume that firm $k \in \mathcal{K}$ wins the share $(\sum_{i\in\mathcal{M}} x^k_i)/(\sum_{i\in\mathcal{M}} \bar{x}_i)$ of the total number of disloyal consumers over any given time-space range $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$, while firm $-k$ wins the remaining disloyal consumers. Thus, the more capacity a firm has, the more disloyal consumers she wins. Then we can describe the rental demand for each firm $k \in \mathcal{K}$ as follows:

$$\overline{w}_e = Y^0_{i,i',t,t'} \frac{\sum_{i\in\mathcal{M}} x^k_i}{\sum_{i\in\mathcal{M}} \bar{x}_i} + Y^k_{i,i',t,t'},\ \forall\, e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}},\ k \in \mathcal{K}. \tag{2}$$

Equations (2) reflect the impact of vehicle accessibility and availability on consumer demands.

Second, for each arc $e \in \mathcal{E}$, we let w_e denote its realized flow. Given any nonnegative integers a and b, we let $[a, b]_{\mathbb{Z}}$ denote the set of all integers between a and b; that is, $[a, b]_{\mathbb{Z}} = \{a, a+1, \ldots, b\}$ if $a \leq b$, and $[a, b]_{\mathbb{Z}} = \emptyset$ if $a > b$. Thus, given any node $n^k_{i,t} \in \mathcal{N}$, the number of micromobility vehicles flowing into this node should be the same as the number of vehicles flowing out from this node:

$$\sum_{e\in f^+(n^k_{i,0})} w_e - x^k_i = 0,\ \forall\, i \in \mathcal{M},\ k \in \mathcal{K}, \tag{3}$$

$$\sum_{e\in f^+(n^k_{i,t})} w_e - \sum_{e\in f^-(n^k_{i,t})} w_e = 0,\ \forall\, t \in [1, T-1]_{\mathbb{Z}},\ i \in \mathcal{M},\ k \in \mathcal{K}, \tag{4}$$

$$-\sum_{e\in f^-(n^k_{i,T})} w_e + x^k_i = 0,\ \forall\, i \in \mathcal{M},\ k \in \mathcal{K}, \tag{5}$$

where $f^+(n^k_{i,t})$ and $f^-(n^k_{i,t})$ denote the sets of arcs that originate and terminate at node $n^k_{i,t}$, respectively. Constraints (3) state that in period $t = 0$, the *outflow* of node $n^k_{i,0}$ equals the number of vehicles initially allocated in region $i \in \mathcal{M}$ by firm $k \in \mathcal{K}$. Constraints (5) state that the number of vehicles of firm $k \in \mathcal{K}$ in region $i \in \mathcal{M}$ and period T (i.e., the end of the operational horizon) returns back to the original status in period 0.

Finally, we have the following three constraints:

$$w_e \leq \overline{w}_e, \ \forall\, e \in \mathcal{E}^k_{\mathrm{R}},\ k \in \mathcal{K}, \tag{6}$$

$$w_e \leq x^k_i, \ \forall\, e \in \mathcal{E}^k_{\mathrm{I}},\ k \in \mathcal{K}, \tag{7}$$

$$w_e \geq 0, \ \forall\, e \in \mathcal{E}^k,\ k \in \mathcal{K}. \tag{8}$$

Constraints (6) ensure that the number of vehicles fulfilling the demand on each rental arc is no larger than the number of rental requests. Constraints (7) ensure that the number of idling vehicles on each arc is no larger than the number of vehicles allocated in region $i \in \mathcal{M}$; that is, each firm is not allowed to park too many vehicles in each region, thereby avoiding traffic congestion. Meanwhile, constraints (7) strengthen the impacts of the initial allocation decisions on the subsequent operation decisions, thereby encouraging the operator to make allocation decisions with greater caution and leading to more reliable allocation decisions. In addition, as adding constraints (7) ensures the generality of our proposed model, one may remove such constraints if they are not needed. Constraints (8) ensure that the realized flow on each arc is non-negative.

In the operation stage, each firm $k \in \mathcal{K}$ relocates micromobility vehicles with respect to any possible scenario to maximize her expected profit subject to uncertain demands $\overline{w}_e$ for each rental arc $e \in \mathcal{E}^k_{\mathrm{R}}$. We approximate the joint distribution of random variables $(Y^A_{i,i',t,t'}, Y^B_{i,i',t,t'}, Y^0_{i,i',t,t'})$ over any given time-space range $(i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}$ using a set of finite scenarios $\mathcal{S}$. Each scenario $s \in \mathcal{S}$ has a probability $\theta_s \geq 0$ with $\sum_{s\in\mathcal{S}} \theta_s = 1$. For each scenario $s \in \mathcal{S}$, each firm $k \in \mathcal{K}$ makes recourse decisions (e.g., vehicle relocation) for all the periods in $\mathcal{T}$. We reuse the aforementioned notation and add a superscript s to each decision variable in the operation stage for each scenario. For each $s \in \mathcal{S}$ and $k \in \mathcal{K}$, we let $\mathbf{w}^{k,s} = [w^s_e,\ \forall\, e \in \mathcal{E}^k]^\top$ and $W^{k,s}(\mathbf{x}^k) = \{\mathbf{w}^{k,s} \in \mathbb{R}^{|\mathcal{E}^k_{\mathrm{R}}|+|\mathcal{E}^k_{\mathrm{L}}|+|\mathcal{E}^k_{\mathrm{I}}|} \mid (2)-(8)\}$, where $|\mathcal{E}^k_{\mathrm{R}}| = \sum_{i\in\mathcal{M}} \sum_{j\in\mathcal{M}\setminus\{i\}} \max\{0, T-\ell_{i,j}\}$, $|\mathcal{E}^k_{\mathrm{L}}| = \sum_{i\in\mathcal{M}} \sum_{j\in\mathcal{M}\setminus\{i\}} \max\{0, T-r_{i,j}\}$, and $|\mathcal{E}^k_{\mathrm{I}}| = M(T-1)$. Given the initial vehicle allocation $\mathbf{x}^k$ for each firm $k \in \mathcal{K}$, the expected net cost $\varphi^k(\mathbf{x}^k)$ in the operation stage can be determined by solving the following network flow optimization problem:

$$\varphi^k(\mathbf{x}^k) = \min_{\substack{\mathbf{w}^{k,s} \in W^{k,s}(\mathbf{x}^k),\\ \forall\, s\in\mathcal{S}}} \ \sum_{s\in\mathcal{S}} \theta_s \left(\sum_{e\in\mathcal{E}^k_{\mathrm{R}}} \left(h^k_{\mathrm{P}}(\overline{w}^s_e - w^s_e) - h^k_{\mathrm{R}} w^s_e \right) + \sum_{e\in\mathcal{E}^k_{\mathrm{I}}} h^k_{\mathrm{I}} w^s_e + \sum_{e\in\mathcal{E}^k_{\mathrm{L}}} h^k_{\mathrm{L}} w^s_e \right). \tag{$\mathcal{O}^k$}$$

3.5 Equivalent Reformulation

Each firm $k \in \mathcal{K}$ simultaneously solves her decision-making model ($\mathcal{D}^k(\mathbf{x}^{-k})$) to the optimality, and eventually, both firms reach a Nash equilibrium between them. To examine such an equilibrium, we first represent the optimality condition of problem ($\mathcal{D}^k(\mathbf{x}^{-k})$) for each firm $k \in \mathcal{K}$ and then integrate the optimality conditions of both firms to seek a feasible solution $(\mathbf{x}^A, \mathbf{x}^B)$ from the integrated model that satisfies both conditions. It follows that firm k's vehicle allocation strategy $\mathbf{x}^k$ must be the best response to the other firm's strategy $\mathbf{x}^{-k}$. Note that problem ($\mathcal{D}^k(\mathbf{x}^{-k})$) is a linear program for each firm $k \in \mathcal{K}$. Thus, we represent its optimality condition using the Karush-Kuhn-Tucker (KKT) conditions and use $\mathcal{W}_0$ to denote the feasible region defined by the KKT conditions of model ($\mathcal{D}^k(\mathbf{x}^{-k})$) for any $k \in \mathcal{K}$ (see Appendix B for details).

As multiple equilibria may exist, we can adopt a certain criterion to choose an equilibrium that satisfies such a criterion. In this paper, we consider two criteria. First, both the regulator and the firms care about the demand loss. On the one hand, the regulator hopes to see more residents use shared micromobility services, thereby supporting a highly convenient and sustainable society. On the other hand, shared micromobility firms hope to satisfy more consumers, thereby generating higher revenues. Thus, while minimizing the net cost (i.e., maximizing the profit) in Problem ($\mathcal{D}^k(\mathbf{x}^{-k})$), each firm $k \in \mathcal{K}$ would prefer an equilibrium strategy that minimizes the expected total demand loss in the market. That is, we solve the following optimization problem:

$$\Gamma_0^{\mathrm{DL}} = \min_{\mathbf{\Lambda}_0 \in \mathcal{W}_0} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_s \sum_{e \in \mathcal{E}_{\mathrm{R}}^k} \left(\overline{w}_e^s - w_e^s \right), \tag{9}$$

where $\mathbf{\Lambda}_0$ denotes the vector of all variables associated with both firms in $\mathcal{K}$ (see Appendix C).

Second, both the regulator and the firms care about the number of allocated micromobility vehicles. On the one hand, the regulator would not prefer too many vehicles, which may cause traffic chaos. On the other hand, some shared micromobility firms, including non-profit organizations, may be concerned about both surviving in the industry and balancing between reducing traffic and serving the public. Thus, each firm $k \in \mathcal{K}$ would prefer an equilibrium strategy that minimizes the total number of allocated vehicles in the city while maintaining a certain service level. Specifically, we consider the following optimization problem:

$$\Gamma_0^{\text{TA}} = \min_{\mathbf{\Lambda}_0 \in \mathcal{W}_0} \left\{ \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{M}} x_i^k \;\middle|\; \sum_{e \in \mathcal{E}_{\text{R}}^k} w_e^s \geq \epsilon \sum_{e \in \mathcal{E}_{\text{R}}^k} \overline{w}_e^s,\ \forall\, k \in \mathcal{K},\ s \in \mathcal{S} \right\}, \tag{10}$$

by which we eventually choose an equilibrium such that the total number of allocated vehicles is minimized and each firm $k \in \mathcal{K}$ guarantees a minimum service level $\epsilon \in [0, 1]$.

3.6 Capacity Sharing

We further extend the model ($\mathcal{D}^k(\mathbf{x}^{-k})$) for each firm $k \in \mathcal{K}$ by considering capacity sharing between the two firms in $\mathcal{K}$. Under a capacity-sharing agreement, each firm $k \in \mathcal{K}$ can share her spare vehicles with firm $-k$ if the latter is in shortage of vehicles to satisfy her demands. To prevent the abuse of this agreement (e.g., a firm does not build any capacity), the firms are required to return the borrowed vehicles within a period of $\Delta = \max\{\ell_{ij} \mid \forall, i, j \in \mathcal{M},\ i \neq j\}$. Thus, in addition to the operational decisions of each firm $k \in \mathcal{K}$ in the problem ($\mathcal{O}^k$), the firm k makes two more decisions in each period and service region: (i) the number of vehicles that the firm shares with her opponent (i.e., firm k sends her own vehicles to firm $-k$); (ii) the number of vehicles that the firm returns to her opponent (i.e., firm k returns firm $-k$'s vehicles back).

We expand the spatial-temporal network $\mathcal{G}$ by adding two types of arcs to the set $\mathcal{E}$ (see Figure 2): (i) *Transfer arcs*: The flow on each transfer arc $e = (n_{i,t}^k, n_{i,t}^{-k}) \in \mathcal{E}_{\text{T}}^k$ represents the number of vehicles transferred from firm $k \in \mathcal{K}$ to firm $-k$ in region $i \in \mathcal{M}$ in period $t \in \mathcal{T}$ to satisfy the consumer demands of firm $-k$, by which firm $-k$ pays firm k a compensation $h_{\text{T}} \geq 0$ for each vehicle. (ii) *Return arcs*: The flow on each return arc $e = (n_{i,t}^k, n_{i,t}^{-k}) \in \mathcal{E}_{\text{N}_\delta}^k$ for any $\delta \in \{1, 2, \ldots, \Delta\}$ represents the number of vehicles that firm k receives in period $t \in [1, T-1]_{\mathbb{Z}}$ and returns back to region $i \in \mathcal{M}$ in period $t + \delta \leq T$. No cost is incurred in the return arcs. Thus, the set $\mathcal{A}$ in Section 3.2 is expanded to $\{\text{R}, \text{I}, \text{L}, \text{T}, \text{N}_\delta,\ \forall\, \delta = 1, 2, \ldots, \Delta\}$, by which $\mathcal{E}^k = \cup_{\forall\, a \in \mathcal{A}} \mathcal{E}_a^k$ is also updated. Additionally, for each $k \in \mathcal{K}$, we let $\mathbf{w}_{\text{C}}^k = [w_e^s,\ \forall\, e \in \mathcal{E}_{\text{T}}^k \cup \mathcal{E}_{\text{N}_1}^k \cup \ldots \cup \mathcal{E}_{\text{N}_\Delta}^k, s \in \mathcal{S}]^\top$ denote the vector of realized flows on all transfer arcs and return arcs from firm k to $-k$ in all scenarios.

With the expanded network $\mathcal{G}$ and settings above, the two-stage stochastic program-

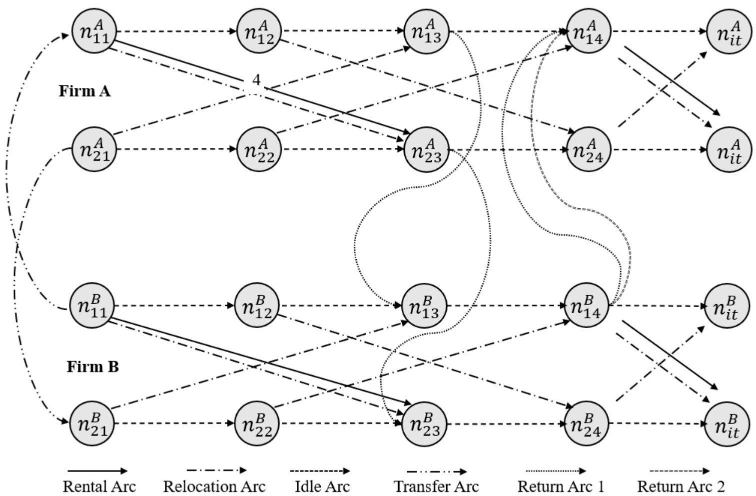


Figure 2. Expanded Spatial-Temporal Network $\mathcal{G}$

ming model for each firm $k \in \mathcal{K}$ is updated as follows:

$$
\begin{aligned}
-\Psi^k = \min\ \mathbf{c}^\top \mathbf{x}^k + \sum_{s\in\mathcal{S}} \theta_s \Bigg(& \sum_{e\in\mathcal{E}_\mathrm{R}^k} \left(h_\mathrm{P}^k(\overline{w}_e^s - w_e^s) - h_\mathrm{R}^k w_e^s \right) \qquad (\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}_\mathrm{C}^{-k})) \\
& + \sum_{e\in\mathcal{E}_\mathrm{I}^k} h_\mathrm{I}^k w_e^s + \sum_{e\in\mathcal{E}_\mathrm{L}^k} h_\mathrm{L}^k w_e^s - \sum_{e\in\mathcal{E}_\mathrm{T}^k} h_\mathrm{T} w_e^s + \sum_{e\in\mathcal{E}_\mathrm{T}^{-k}} h_\mathrm{T} w_e^s \Bigg)
\end{aligned}
$$

s.t. (1); (2) − (4), (6) − (8), $\forall\, s \in \mathcal{S}$,

$$
\sum_{e\in f^+(n_{i,T}^k)} w_e^s - \sum_{e\in f^-(n_{i,T}^k)} w_e^s + x_i^k = 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ s \in \mathcal{S}, \tag{11}
$$

$$
\sum_{i\in\mathcal{M}} \sum_{e\in\mathcal{E}_\mathrm{T}^{-k}(t)} w_e^s - \sum_{i\in\mathcal{M}} \sum_{\delta=1}^{\min\{\Delta, T-t\}} \sum_{e\in\mathcal{E}_{\mathrm{N}_\delta}^k(t)} w_e^s = 0,\ \forall\, k \in \mathcal{K},\ t \in [1, T-1]_\mathbb{Z},\ s \in \mathcal{S}, \tag{12}
$$

where $\mathcal{E}_\mathrm{T}^k(t) = \{(n_{i,t}^k, n_{i,t}^{-k}) \in \mathcal{E}_\mathrm{T}^k \mid \forall\, i \in \mathcal{M}\}$ and $\mathcal{E}_{\mathrm{N}_\delta}^k(t) = \{(n_{i,t}^k, n_{i,t}^{-k}) \in \mathcal{E}_{\mathrm{N}_\delta}^k \mid \forall\, i \in \mathcal{M}\}$ for any $t \in \mathcal{T}$, $k \in \mathcal{K}$, and $\delta \in [1, \min\{\Delta, T-t\}]$, constraints (11) are updated from (5) because of the inclusion of return arcs in $\mathcal{G}$, and constraints (12) ensure that all the vehicles that firm $k \in \mathcal{K}$ borrows from firm $-k$ should be returned back to firm $-k$. Note that it is easy for both firms to comply with the proposed capacity-sharing agreement because each firm can easily receive idling vehicles from her opponent and return them to the opponent in any service region.

We can then examine the Nash equilibrium of the game between Firm A and Firm

B under the capacity-sharing agreement by formulating the KKT conditions of model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}_{\mathrm{C}}^{-k})$) for any $k \in \mathcal{K}$, where the feasible region defined by the KKT conditions is denoted by $\mathcal{W}_1$ (see Appendix D for details). We also use the two criteria in Section 3.5 to select the corresponding Nash equilibria, respectively. For the first criterion concerning the expected total demand loss, we solve the following optimization problem:

$$\Gamma_1^{\mathrm{DL}} = \min_{\mathbf{\Lambda}_1 \in \mathcal{W}_1} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_s \sum_{e \in \mathcal{E}_{\mathrm{R}}^k} \left(\overline{w}_e^s - w_e^s \right), \tag{13}$$

where $\mathbf{\Lambda}_1$ denotes the vector of all variables associated with both firms in $\mathcal{K}$ (see Appendix E for details). For the second criterion concerning the total number of allocated micromobility vehicles, we solve the following optimization problem:

$$\Gamma_1^{\mathrm{TA}} = \min_{\mathbf{\Lambda}_1 \in \mathcal{W}_1} \left\{ \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{M}} x_i^k \;\middle|\; \sum_{e \in \mathcal{E}_{\mathrm{R}}^k} w_e^s \geq \epsilon \sum_{e \in \mathcal{E}_{\mathrm{R}}^k} \overline{w}_e^s, \ \forall\, k \in \mathcal{K},\ s \in \mathcal{S} \right\}. \tag{14}$$

The following proposition shows that the capacity-sharing agreement between both firms in $\mathcal{K}$ benefits the entire system by reducing the expected total demand loss.

Proposition 1 *When Γ_0^{DL} exists, we have $\Gamma_1^{\mathrm{DL}} - \Gamma_0^{\mathrm{DL}} \leq 0$.*

In addition to models (13) and (14) that consider loyal and disloyal consumers together, we provide alternative models that consider them separately. However, despite offering more targeted services, these models perform less effectively than those considering loyal and disloyal consumers together (see Appendix G for details).

Remark 1 *We can alternatively determine the share of disloyal consumers that firm $k \in \mathcal{K}$ can attract in region $i \in \mathcal{M}$ and period $t \in \mathcal{T}$ based on her available vehicles in this region and period, slightly different from how we determine the share of disloyal consumers via equations* (2)*. Through numerical experiments, we find that the alternative equations considering available vehicles exhibit performance similar to equations* (2)*, but with greater computational complexity (see Appendix H for details).*

4 Numerical Experiments: A Case Study

We conduct numerical experiments based on real data from Citi Bike (2022). We first discuss the parameter settings and then obtain managerial insights from various experiments.

All the numerical experiments are performed on a computing node with 24 2.3-GHz Intel Xeon E5-2670 processors and 32 GB of memory in a high-performance computing cluster. IBM ILOG CPLEX 12.10, under its default setting, is used as the mixed-integer programming (MIP) solver.

4.1 Parameter Settings

We collect data from Citi Bike (2022) in New York City (NYC) from January 1 to December 31, 2019. We focus on Midtown Manhattan in a rectangular area formed by four intersecting streets in NYC: First Avenue, Eleventh Avenue, Twenty-third Street, and Fifty-seventh Street (see Figure 3), and obtain 2,320,205 trips in this area. We specify the geographical locations of all bike docking stations in Figure 3 and divide the area into six service regions (i.e., $M = 6$) using k-means clustering based on the Euclidean distance between any two stations. Two stations that are close to each other can be clustered together. The dots of different sizes indicate the station locations and their demand levels, i.e., a larger dot indicates a higher demand level. Note that considering six service regions helps us retain practical features in our models while not leading to a large-scale and expanded spatial-temporal network $\mathcal{G}$, under which the models are difficult to solve.

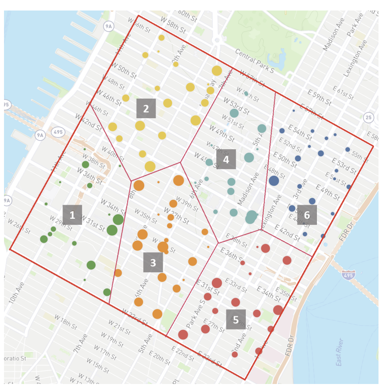


Figure 3. Service Regions in Midtown Manhattan

Table 1. Trip Duration Across Service Regions

Regions	1	2	3	4	5	6
1	—	1	1	2	1	2
2	1	—	1	1	2	1
3	1	1	—	1	1	1
4	2	1	1	—	1	1
5	1	2	1	1	—	1
6	2	1	1	1	1	—

Based on the trip data, we find that the average trip duration from one region to a neighboring region is 9.77 minutes. For simplicity, we assume the traveling speed between any two neighboring regions is the same and set each period as 10 minutes (leading to 144 periods per day). Table 1 summarizes the trip duration in terms of the number of periods (i.e., ℓ_{ij}) between any two service regions $i, j \in \mathcal{M}$ and $i \neq j$.

Figure 4 shows the average number of consumer trips in the entire area in a day, with the consumer trips between any starting and destination regions summarized in Figure 18 (see Appendix I). We note that most trip demands occur during the daytime, especially from early morning (approximately 06:00) until night (approximately 22:00), while the demand from 22:00–6:00 is almost negligible. We also note that the demand pattern in the late afternoon exhibits similarity with that in the early morning in most of the regions. Hence, hereafter we assume that the operational horizon for each firm $k \in \mathcal{K}$ is 8 hours, from 6:00 to 14:00 (i.e., the shaded area in Figure 4), leading to $T = 48$ periods in our models. It is worth noting that there is at least one peak in one day, i.e., high volumes of trip demands, regardless of the starting and destination regions. Because the range from 6:00–14:00 covers at least one peak for any starting and destination regions, this operational horizon is appropriate enough to capture both the demand trend and traffic peak.

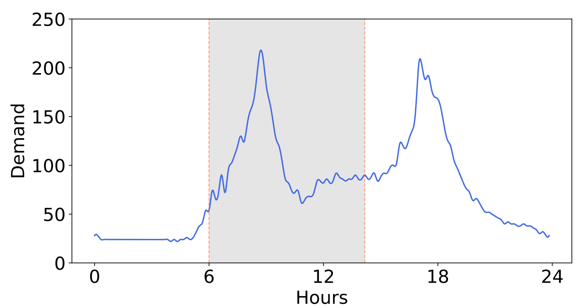
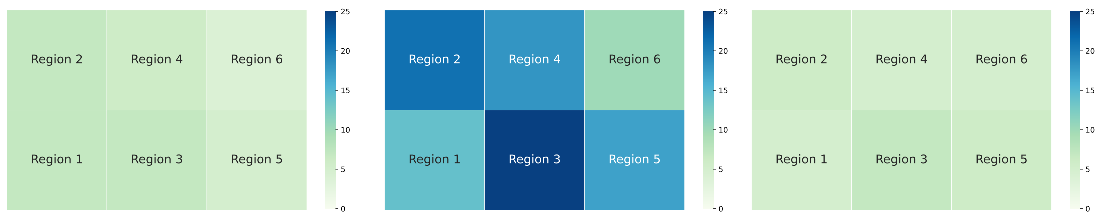


Figure 4. Average Number of Consumer Trips in a Day in Midtown Manhattan

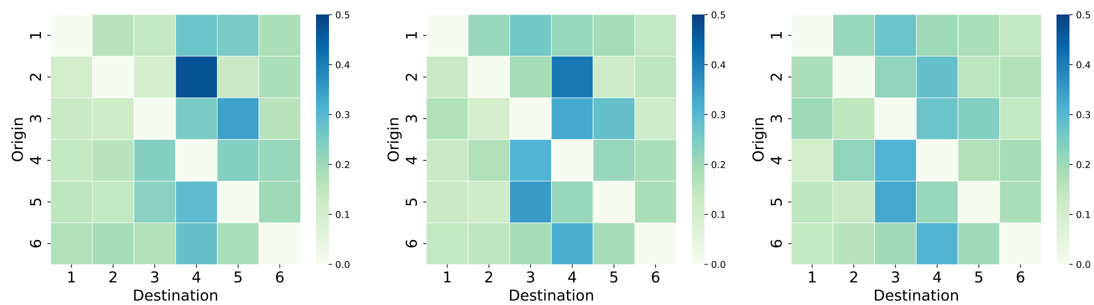
For any given 3-tuple (*starting region*, *destination region*, *starting period*), we use the data to obtain the average number of trips (i.e., demands) over all the given data samples between the given starting and destination regions in the given starting period. We let $\overline{D}_{i,j,t}$ denote the average demands from region $i \in \mathcal{M}$ to region $j \in \mathcal{M} \setminus \{i\}$ in period $t \in \mathcal{T}$ (see Figure 18 in Appendix I for a summary of all these average demands). We let $\hat{D}_{i,t} = \sum_{j \in \mathcal{M} \setminus \{i\}} \overline{D}_{i,j,t}$ for any $i \in \mathcal{M}$ and $t \in \mathcal{T}$, representing the total number of trips originating from region i in period t. Figure 5 illustrates the total number of trips originating from each region in three periods: $t = 6$ (low demand level), $t = 18$ (high demand level), and $t = 36$ (low demand level). Clearly, the demands in $t = 18$ represent a peak, and regions 2 and 3 are the busiest. We also let $\phi_{i,t,j} = \overline{D}_{i,j,t} / \hat{D}_{i,t}$ for any $i \in \mathcal{M}$, $j \in \mathcal{M} \setminus \{i\}$, and $t \in \mathcal{T}$, denoting the percentage of all the trips originating from region i in period t that eventually go to region j. Figure 6 illustrates the values of $\phi_{i,t,j}$ for any $i \in \mathcal{M}$ and

$j \in \mathcal{M} \setminus \{i\}$ when $t \in \{6, 18, 36\}$.



(a) $t = 6$ (7:00–7:10) **(b)** $t = 18$ (9:00–9:10) **(c)** $t = 36$ (12:00–12:10)

Figure 5. The Number of Trips Originating From Each Region



(a) $t = 6$ (7:00–7:10) **(b)** $t = 18$ (9:00–9:10) **(c)** $t = 36$ (12:00–12:10)

Figure 6. Trip Percentages

For each firm $k \in \mathcal{K}$, we estimate her cost parameters in USD as follows: The micro-mobility vehicle allocation cost[1] for any $i \in \mathcal{M}$ is $c_i^k = 0.16$. The revenue per vehicle trip[2] is $h_{\mathrm{R}}^k = 0.4$ and the penalty cost per unsatisfied demand[3] is $h_{\mathrm{P}}^k = 0.1$. The relocation cost per relocated vehicle[4] is $h_{\mathrm{L}}^k = 0.12$ and the idle cost per idling vehicle[5] is $h_{\mathrm{I}}^k = 0.06$. The transfer cost per vehicle[6] is $h_{\mathrm{T}} = 0.12$. Given a budget limit to allocate vehicles in $\mathcal{M}$, we set $\hat{x}^k = 1500$, which is sufficiently large because the number of vehicles that each firm eventually allocates is less than $\hat{x}^k$.

The upper bound $\bar{x}_i$ of the total number of vehicles allocated by the two firms in $\mathcal{K}$ in each region $i \in \mathcal{M}$, i.e., *initial vehicle allocation quota*, is not available in the data. Nevertheless, we have $\max\{\sum_{i \in \mathcal{M}} \hat{D}_{i,t} \mid \forall t \in \mathcal{T}\} = 358$. To have a proper number of vehicles in the area to satisfy the trip demands of both firms while not creating traffic chaos on the street, we set $\sum_{i \in \mathcal{M}} \bar{x}_i = 360$. Given the total number of trips originating from region i in period t, $\hat{D}_{i,t}$, for any $i \in \mathcal{M}$ and $t \in \mathcal{T}$, we further set $\bar{x}_i = 360 \times (\sum_{t \in \mathcal{T}} \hat{D}_{i,t}) / (\sum_{i \in \mathcal{M}} \sum_{t \in \mathcal{T}} \hat{D}_{i,t})$ for any $i \in \mathcal{M}$. For any two regions $i, i' \in \mathcal{M}$ and $i \neq i'$, we set $r_{i,i'} = \ell_{i,i'}$.

We further generate scenarios in $\mathcal{S}$. We use a random parameter $\tilde{D}_{i,t}$ to represent the number of trips (i.e., market demands) originating from region $i \in \mathcal{M}$ in period $t \in \mathcal{T}$. We assume that $(\tilde{D}_{1,t}, \ldots, \tilde{D}_{M,t})^\top$ follows a multivariate normal distribution; that is, $(\tilde{D}_{1,t}, \ldots, \tilde{D}_{M,t})^\top \sim \mathcal{N}(\hat{\boldsymbol{D}}_t, \boldsymbol{\Sigma}_t)$, for any $t \in \mathcal{T}$, where $\hat{\boldsymbol{D}}_t = (\hat{D}_{1,t}, \ldots, \hat{D}_{M,t})^\top$ and $\boldsymbol{\Sigma}_t$ represents the corresponding covariance matrix estimated from the given data in each period t. Following this distribution, we randomly generate $1,000$ samples of trip demands and use k-means clustering to cluster them into three groups, where we use the center of each group to represent a demand scenario in $\mathcal{S}$ and thus set $|\mathcal{S}| = 3$. Note that we do not choose a large number of scenarios for two reasons. First, as our models, i.e., (9)–(10) and (13)–(14), are large-scale MIPs that can be difficult to solve, our model will become computationally intractable if using a large number of scenarios. Second, the real data shows an apparent demand trend; thus, a few scenarios can well represent demand uncertainty. We also discuss the impact of $|\mathcal{S}|$ on computational performance in detail in Appendix J, thereby demonstrating that setting $|\mathcal{S}| = 3$ is effective enough.

Now, given the market demand originating from region $i \in \mathcal{M}$ in period $t \in \mathcal{T}$ and scenario $s \in \mathcal{S}$, denoted by $\dot{D}_{i,t,s}$, we let $D_{i,i',t,t',s} = \dot{D}_{i,t,s}\phi_{i,t,i'}$ for any $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$ and $s \in \mathcal{S}$, representing the trip demand from region i in period t to region i' in period $t' = t + \ell_{i,i'}$ in scenario $s \in \mathcal{S}$. Then, given a percentage (α_k) of consumers who are loyal to each firm $k \in \mathcal{K}$, we have the number of firm k's loyal consumers $Y^k_{i,i',t,t',s} = \alpha_k D_{i,i',t,t',s}$ and the number of disloyal consumers $Y^0_{i,i',t,t',s} = (1 - \alpha_A - \alpha_B) D_{i,i',t,t',s}$ over any $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$ and in any scenario $s \in \mathcal{S}$.

4.2 Impact of Initial Vehicle Allocation Quota

We examine the impact of the initial vehicle allocation quota on the firms' performance under the parameter settings in Section 4.1. We vary the quota for the entire area $\sum_{i\in\mathcal{M}} \bar{x}_i$ and the percentages of the two firms' loyal consumers (α_A, α_B), while keeping other parameters unchanged and not adopting capacity sharing. With $\sum_{i\in\mathcal{M}} \bar{x}_i$ given at 360, representing a proper-quota case, we also consider $\sum_{i\in\mathcal{M}} \bar{x}_i$ takes half and double of this quota, leading to 180 and 720, representing the low-quota and large-quota cases, respectively. When $\sum_{i\in\mathcal{M}} \bar{x}_i$ is given, we let $\bar{x}_i = \sum_{i\in\mathcal{M}} \bar{x}_i \times (\sum_{t\in\mathcal{T}} \hat{D}_{i,t}) / (\sum_{i\in\mathcal{M}} \sum_{t\in\mathcal{T}} \hat{D}_{i,t})$ for any $i \in \mathcal{M}$.

First, we consider the two firms in $\mathcal{K}$ are symmetric with respect to (α_A, α_B), which can be $(0, 0)$, $(0.25, 0.25)$, and $(0.45, 0.45)$. Given $\bar{x}_i$ for any $i \in \mathcal{M}$ and (α_A, α_B), we solve Problem (9) (that minimizes the expected total demand loss) and Problem (10) (that minimizes the total number of allocated micromobility vehicles). Table 2 shows the equilibrium results. For each instance, we use a 3-tuple (*Firm A's result*, *Firm B's result*, *Summation of both firms' results*) to show both firms' performance in the initial vehicle allocation, profit, and service level.

Table 2. Impact of Initial Vehicle Allocation Quota on Symmetric Firms

	(α_A, α_B)	**Allocation**		**Profit**		**Service Level (%)**	
		Min. Demand Loss	**Min. Allocation**	**Min. Demand Loss**	**Min. Allocation**	**Min. Demand Loss**	**Min. Allocation**
Low (180)	(0, 0)	(0, 180, 180)	(0, 180, 180)	(0, 1801.5, 1801.5)	(0, 1801.5, 1801.5)	(-, 94.3, 94.3)	(-, 94.3, 94.3)
	(0.25, 0.25)	(90, 90, 180)	(90, 90, 180)	(900.3, 900.3, 1,800.6)	(900.3, 900.3, 1,800.6)	(94.3, 94.3, 94.3)	(94.3, 94.3, 94.3)
	(0.45, 0.45)	(90, 90, 180)	(90, 90, 180)	(900.2, 900.2, 1,800.4)	(900.2, 900.2, 1,800.4)	(94.3, 94.3, 94.3)	(94.3, 94.3, 94.3)
Proper (360)	(0, 0)	(0, 360, 360)	(0, 360, 360)	(0, 1,415.3, 1,415.3)	(0, 1,415.3, 1,415.3)	(-, 99.7, 99.7)	(-, 99.7, 99.7)
	(0.25, 0.25)	(180, 180, 360)	(180, 180, 360)	(707.7, 707.7, 1,415.4)	(707.7, 707.7, 1,415.4)	(99.7, 99.7, 99.7)	(99.7, 99.7, 99.7)
	(0.45, 0.45)	(180, 180, 360)	(180, 180, 360)	(707.7, 707.7, 1,415.4)	(707.7, 707.7, 1,415.4)	(99.7, 99.7, 99.7)	(99.7, 99.7, 99.7)
Large (720)	(0, 0)	(0, 720, 720)	(0, 720, 720)	(0, 329.4, 329.4)	(0, 329.4, 329.4)	(-, 100, 100)	(-, 100, 100)
	(0.25, 0.25)	(360, 360, 720)	(65, 65, 130)	(164.6, 164.6, 329.2)	(521.7, 521.7, 1043.4)	(100, 100, 100)	(97.3, 97.3, 97.3)
	(0.45, 0.45)	(360, 360, 720)	(83, 83, 166)	(164.6, 164.6, 329.2)	(830.9, 830.9, 1,661.8)	(100, 100, 100)	(94.2, 94.2, 94.2)

Table 2 suggests that setting a proper quota approximating trip demands can help the firms achieve a high service level and profit simultaneously. Specifically, compared to the low-quota case, the service level of each firm is increased by around 5% under the proper-quota case regardless of the equilibrium criterion and loyal consumer percentages. Compared to the large-quota case, the proper-quota case performs better under different cases: (i) the profit of each firm is increased regardless of the equilibrium criterion when $\alpha_A \in \{0, 0.25\}$ and under the criterion of demand loss when $\alpha_A = 0.45$; (ii) the service level of each firm is also increased under the criterion of initial allocation when $\alpha_A \in \{0.25, 0.45\}$.

Both firms use up the quota under the low- and proper-quota cases regardless of the equilibrium criterion and loyal consumer percentages. For such cases, the number of disloyal consumers each firm can attract by allocating one more vehicle, i.e., $1/\sum_{i\in\mathcal{M}} \overline{x}_i$, is relatively large, motivating the firms to allocate as many vehicles as possible to attract disloyal consumers. Under the large-quota case with a large loyal consumer percentage (i.e., $\alpha_A \in \{0.25, 0.45\}$), both firms use up the quota under the criterion of demand loss and use the quota partially under the other criterion. When the initial vehicle allocation quota is large, $1/\sum_{i\in\mathcal{M}} \overline{x}_i$ is relatively small, discouraging firms from allocating a large number of vehicles to attract disloyal consumers. Thus, when we aim to find an equilibrium minimizing the total number of allocated vehicles, we may not use up the quota, while a large number of vehicles may be required to achieve an equilibrium minimizing the expected total demand loss.

Table 3. Impact of Initial Vehicle Allocation Quota on Asymmetric Firms

	(α_A, α_B)	**Allocation**		**Profit**		**Service Level (%)**	
		Min. Demand Loss	**Min. Allocation**	**Min. Demand Loss**	**Min. Allocation**	**Min. Demand Loss**	**Min. Allocation**
Low (180)	(0.1, 0.15)	(72, 108, 180)	(72, 108, 180)	(719.5, 1,082.0, 1,801.5)	(719.5, 1,082.0, 1,801.5)	(94.3, 94.3, 94.3)	(94.3, 94.3, 94.3)
	(0.1, 0.45)	(33, 147, 180)	(49, 131, 180)	(327.3, 1,473.9, 1,801.2)	(386.2, 1,386.9, 1,773.1)	(94.3, 94.3, 94.3)	(96.3, 92.5, 93.3)
	(0.1, 0.8)	(20, 160, 180)	(20, 160, 180)	(200.0, 1,601.4, 1801.4)	(200.0, 1,601.4, 1801.4)	(94.2, 94.3, 94.3)	(94.2, 94.3, 94.3)
	(0.3, 0.4)	(76, 104, 180)	(80, 100, 180)	(771.5, 1,030.0, 1,801.5)	(763.3, 1,000.7, 1,764.0)	(94.3, 94.3, 94.3)	(93.5, 92.8, 93.1)
	(0.3, 0.6)	(60, 120, 180)	(65,115,180)	(600.2, 1,201.3, 1,801.5)	(602.2, 1,195.2, 1,797.4)	(94.3, 94.3, 94.3)	(95.2, 93.6, 94.1)
Proper (360)	(0.1, 0.15)	(151, 209, 360)	(103, 257, 360)	(582.2, 833.1, 1,415.3)	(471.7, 939.2, 1,410.9)	(99.7, 99.7, 99.7)	(99.2, 99.7, 99.5)
	(0.1, 0.45)	(69, 291, 360)	(55, 305, 360)	(258.1, 1157.2, 1,415.3)	(251.4, 1,159.9, 1,411.3)	(99.8, 99.7, 99.7)	(98.9, 99.7, 99.6)
	(0.1, 0.8)	(39, 321, 360)	(20, 161, 181)	(158.4, 1256.7, 1,415.1)	(189.9, 1,519.1, 1,709.0)	(99.7, 99.7, 99.7)	(95.2, 95.2, 95.2)
	(0.3, 0.4)	(156, 204, 360)	(87, 116, 203)	(605.0, 810.2, 1,415.2)	(651.0, 868.0, 1,519.0)	(99.7, 99.7, 99.7)	(97.9, 97.9, 97.9)
	(0.3, 0.6)	(131, 229, 360)	(60, 121, 181)	(447.8, 967.5, 1,415.3)	(569.7, 1,139.3, 1709.0)	(99.9, 99.6, 99.7)	(95.2, 95.2, 95.2)
Large (720)	(0.1, 0.15)	(343,377,720)	(35, 53, 88)	(108.7, 219.4, 328.1)	(231.9, 347.8, 579.7)	(100, 100, 100)	(98.7, 98.7, 98.7)
	(0.1, 0.45)	(77, 643, 720)	(26, 116, 142)	(137.8, 190.4, 328.3)	(205.1, 923.1, 1,128.2)	(100, 100, 100)	(97.4, 97.4, 97.4)
	(0.1, 0.8)	(43, 677, 720)	(19, 149, 168)	(134.4, 194.2, 328.6)	(184.9, 1479.1, 1,664.0)	(100, 100, 100)	(94.4, 94.4, 94.4)
	(0.3, 0.4)	(64, 86, 150)	(64, 86, 150)	(586.3, 781.8, 1368.1)	(586.3, 781.8, 1,368.1)	(95.7, 95.7,95.7)	(95.7, 95.7,95.7)
	(0.3, 0.6)	(56, 112, 168)	(56, 112, 168)	(554.6, 1109.3, 1663.9)	(554.6, 1109.3, 1,663.9)	(94.4,94.4,94.4)	(94.4,94.4,94.4)

Next, we consider the two firms in $\mathcal{K}$ are asymmetric with respect to (α_A, α_B), i.e., $\alpha_A \neq \alpha_B$. Table 3 shows the equilibrium results. We obtain the following results that are consistent with the symmetric case in Table 2: (i) under the proper-quota case, each firm can attain a high service level and profit; (ii) under the low-quota case, the firms use up the allocation quota; and (iii) under the large-quota case, the firms may not use up the quota, and the total number of allocated vehicles and profits increase with the loyal consumer percentage. We also obtain results that are inconsistent with the symmetric case. Specifically, under the proper-quota case, both firms do not use up the quota when $\alpha_A + \alpha_B$ is large (e.g., 0.7 and 0.9) under the equilibrium criterion of the initial allocation. This happens mainly because the market competitiveness under the asymmetric case when $\alpha_A + \alpha_B$ is large is weaker than that under the symmetric case, where neither firm dominates the market, and both firms have to halve the market equally. That is, under the symmetric case, both firms allocate as many vehicles as possible to attract disloyal consumers and thus use up the quota. In contrast, when $\alpha_A + \alpha_B$ is large under the asymmetric case, both firms are less willing to compete for limited disloyal consumers.

4.3 Impact of Competition

We examine the impact of competition by comparing the performance of the models with and without competition. Specifically, we consider Problems (9) and (10) as the models with competition. For each firm $k \in \mathcal{K}$, we let $\tilde{x}_i^k = \bar{x}_i \times \alpha_k / (\alpha_A + \alpha_B)$ in region $i \in \mathcal{M}$ and define

$$\min_{\mathbf{x}^k} \left\{ \mathbf{c}^{k\top} \mathbf{x}^k + \varphi^k \left(\mathbf{x}^k \right) \,\middle|\, \sum_{i \in \mathcal{M}} x_i^k \leq \hat{x}^k; \quad x_i^k \leq \tilde{x}_i^k, \, \forall\, i \in \mathcal{M}; \quad x_i^k \geq 0, \, \forall\, i \in \mathcal{M} \right\}$$

as the model without competition. Given an initial vehicle allocation quota and loyal consumer percentages, (i) we solve the above model and show the results in Table 4 (see the columns "W/O Competition"); (ii) we select the equilibrium result that yields the highest profit by Problems (9) and (10) (see the columns "W/ Competition" in Table 4). Note that here we consider the "*highest profit*" because the model without competition minimizes the total cost, i.e., maximizes the total profit, ensuring fair comparisons.

Table 4. Impact of Competition Under Asymmetric Cases

	(α_A, α_B)	**Allocation**		**Profit**		**Service Level (%)**	
		W/ Competition	**W/O Competition**	**W/ Competition**	**W/O Competition**	**W/ Competition**	**W/O Competition**
	(0.1, 0.15)	(72, 108, 180)	(70, 110, 180)	(719.5, 1,082.0, 1,801.5)	(705.7, 1095.8, 1801.5)	(94.3, 94.3, 94.3)	(94.1, 94.3, 94.2)
	(0.1, 0.45)	(33, 147, 180)	(31, 149, 180)	(327.3, 1,473.9, 1,801.2)	(319.6, 1481.6, 1801.2)	(94.3, 94.3, 94.3)	(93.7, 94.4, 94.2)
Low	(0.1, 0.8)	(20, 160, 180)	(17, 163, 180)	(200.0, 1,601.4, 1801.4)	(195.3, 1603.8, 1799.1)	(94.2, 94.3, 94.3)	(91.2,94.5,94.2)
(180)	(0.3, 0.4)	(76, 104, 180)	(74, 106, 180)	(771.5, 1,030.0, 1,801.5)	(762.4, 1038.3, 1800.7)	(94.3, 94.3, 94.3)	(93.7, 94.6, 94.2)
	(0.3, 0.6)	(60, 120, 180)	(59, 121,180)	(600.2, 1,201.3, 1,801.5)	(599.5, 1201.9, 1801.4)	(94.3, 94.3, 94.3)	(94.0, 94.4, 94.3)
	(0.1, 0.15)	(151, 209, 360)	(142, 218, 360)	(582.2, 833.1, 1,415.3)	(561.7, 853.7, 1415.3)	(99.7, 99.7, 99.7)	(99.7,99.7,99.7)
	(0.1, 0.45)	(69, 291, 360)	(64, 296, 360)	(258.1, 1157.2, 1,415.3)	(257.0, 1158.3, 1415.3)	(99.8, 99.7, 99.7)	(99.6, 99.7, 99.7)
Proper	(0.1, 0.8)	(20, 161, 181)	(20, 161, 181)	(189.9, 1519.1, 1709.0)	(189.9, 1519.1, 1709.0)	(95.2, 95.2, 95.2)	(95.2, 95.2, 95.2)
(360)	(0.3, 0.4)	(87, 116, 203)	(87, 116, 203)	(651.0, 868.0, 1519.0)	(651.0, 868.0, 1519.0)	(97.9, 97.9, 97.9)	(97.9, 97.9, 97.9)
	(0.3, 0.6)	(60, 121, 181)	(60, 121, 181)	(569.7, 1,139.3, 1709.0)	(569.7, 1,139.3, 1709.0)	(95.2, 95.2, 95.2)	(95.2, 95.2, 95.2)
	(0.1, 0.15)	(35, 53, 88)	(35, 53, 88)	(231.9, 347.8, 579.7)	(231.9, 347.8, 579.7)	(98.7, 98.7, 98.7)	(98.7, 98.7, 98.7)
	(0.1, 0.45)	(26, 116, 142)	(26, 116, 142)	(205.1, 923.1, 1128.2)	(205.1, 923.1, 1128.2)	(97.4, 97.4, 97.4)	(97.4, 97.4, 97.4)
Large	(0.1, 0.8)	(19, 149, 168)	(19, 149, 168)	(184.9, 1479.1, 1,664.0)	(184.9, 1479.1, 1,664.0)	(94.4, 94.4, 94.4)	(94.4, 94.4, 94.4)
(720)	(0.3, 0.4)	(64, 86, 150)	(64, 86, 150)	(586.3, 781.8, 1,368.1)	(586.3, 781.8, 1,368.1)	(95.7, 95.7,95.7)	(95.7, 95.7,95.7)
	(0.3, 0.6)	(56, 112, 168)	(56, 112, 168)	(554.6, 1109.3, 1663.9)	(554.6, 1109.3, 1663.9)	(94.4,94.4,94.4)	(94.4,94.4,94.4)

When $\alpha_A \neq \alpha_B$, we call the firm with a lower loyal consumer percentage the *weak* firm and the other the *strong* firm. For instance, when $(\alpha_A, \alpha_B) = (0.1, 0.45)$, Firm A is the weak firm and Firm B is the strong firm. Table 4 suggests that the weak firm benefits from the competition, which hurts the strong firm though. When no competition exists, the weak firm is given a small quota to allocate vehicles, while she has the opportunity to use the entire quota to allocate more vehicles and earn a higher profit after the competition is introduced. It follows that the strong firm is left with a smaller quota to allocate vehicles than before. Such changes after introducing the competition are more significant when the disloyal consumer percentage (i.e., $1 - \alpha_A - \alpha_B$) is large under the proper-quota case. This is intuitive because more disloyal consumers lead to a more competitive market, where the competition will have a greater impact on the two firms.

We further analyze how each firm allocates vehicles in each region when the competition is present. For each instance of the model with competition described above, we let x_i^{k*} be the equilibrium allocation decision of firm $k \in \mathcal{K}$ in region $i \in \mathcal{M}$ that achieves the highest profit as mentioned above, and define $x_i^{k*}/(x_i^{k*} + x_i^{-k*})$ as the allocation ratio of firm k in region i. For each firm $k \in \mathcal{K}$, we compare this ratio with $\alpha_k/(\alpha_A + \alpha_B)$, which defines the proportion of quota issued to firm k when there is no competition; a larger allocation ratio indicates that the competition facilitates more allocated vehicles. Specifically, we consider Problem (9) with $\alpha_B = 0.45$ and $\alpha_A \in \{0.1, 0.2, 0.3, 0.4\}$, where Firm A is the weak firm. Figure 7 shows the two firms' allocation ratios in every region (see the gray and white bars), where the red dashed line shows the value of $\alpha_A/(\alpha_A + \alpha_B)$. Thus, when the grey bar of region $i \in \mathcal{M}$ exceeds (i.e., is higher than) the red line, i.e., $x_i^{A*}/(x_i^{A*} + x_i^{B*}) > \alpha_A/(\alpha_A + \alpha_B)$, we say the weak firm (i.e., Firm A) allocates more vehicles in region i after the competition is introduced.

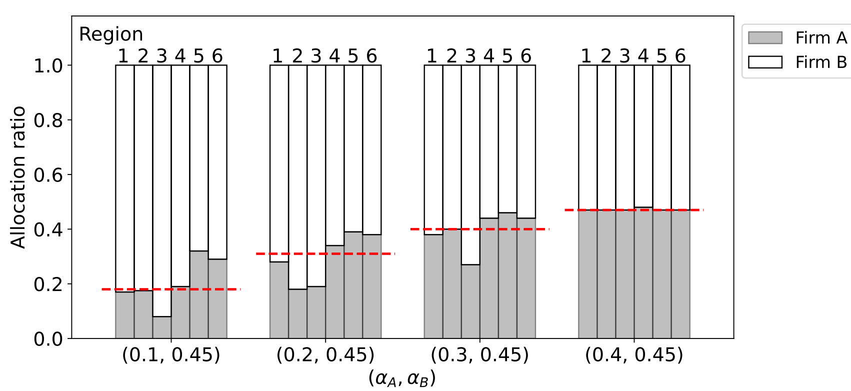


Figure 7. Initial Allocation in Each Region

Figure 7 shows that the weak firm allocates more vehicles in regions 5 and 6 and allocates less in regions 2 and 3 (i.e., the busiest regions with the largest consumer trips), compared to the case without competition. That is, regarding vehicle allocation in each region, the strong firm dominates the busy regions while the weak firm focuses on the non-busy ones. Note that allocating more vehicles in the busy regions is crucial to meet more trip demands, motivating the strong firm to prioritize the vehicle allocation there and allocate less in the non-busy regions. As a result, the weak firm allocates more vehicles to the non-busy regions. As α_A grows, Firm A becomes more comparably competitive with Firm B, which becomes more difficult to allocate vehicles in the busy regions, and eventually, the two firms' performance becomes similar.

4.4 Impact of Capacity Sharing

We investigate the impact of capacity sharing on firms' operations by comparing the results of Problem (9) (without capacity sharing) and Problem (13) (with capacity sharing). First, we consider the two firms in $\mathcal{K}$ are symmetric with respect to (α_A, α_B), i.e., $\alpha_A = \alpha_B \in \{10\% + i \times 5\%, \forall i \in [0,7]_{\mathbb{Z}}\}$. For any given α_A, we obtain the same symmetric equilibrium result after solving Problem (13) (see Table 5). In this equilibrium, neither firm would like to share vehicles with her opponent.

Table 5. Symmetric Equilibrium for Symmetric Firms

Allocation		Profit		Transfer		Demand		Service Level	
Firm A	Firm B	Firm A	Firm B	Firm A	Firm B	Firm A	Firm B	Firm A	Firm B
180	180	707.67	707.67	0	0	2724.04	2724.04	99.70	99.70

For any given α_A, we also obtain an asymmetric equilibrium (see Table 6). Firm A allocates fewer vehicles and has fewer total demands than Firm B. Despite fewer allocated vehicles, Firm A can receive vehicles transferred from Firm B due to the capacity-sharing agreement and thus achieves a high service level. Furthermore, when each firm has more loyal consumers, i.e., α_A is larger, the two firms' performance becomes closer, and the number of vehicles shared between the two firms drops. Thus, even when two firms are symmetric, one firm may act like a free rider that allocates fewer vehicles and asks for vehicles transferred from her opponent if needed.

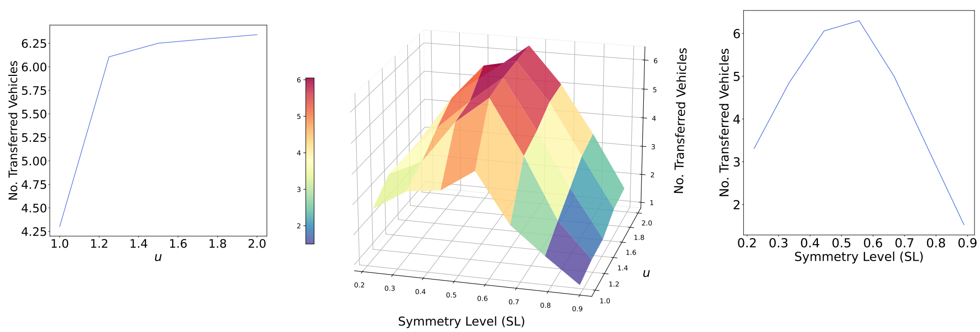
Table 6. Asymmetric Equilibrium for Symmetric Firms

$\alpha_A(\alpha_B)$	Allocation		Profit ($)		Transfer		Demand		Service Level (%)	
	Firm A	Firm B	Firm A	Firm B	Firm A	Firm B	Firm A	Firm B	Firm A	Firm B
10%	74	286	193.35	1221.98	0	4.37	1440.72	4007.37	99.94	99.61
15%	108	252	409.07	1006.26	0	2.54	1961.31	3486.77	99.78	99.65
20%	150	210	554.64	860.69	0	1.95	2457.52	2990.57	99.80	99.61
25%	164	196	626.10	789.23	0	1.32	2601.46	2846.62	99.80	99.60
30%	171	189	650.22	765.11	0	1.02	2668.38	2779.71	99.72	99.67
35%	173	187	668.51	746.82	0	0.71	2693.76	2754.32	99.68	99.71
40%	176	184	687.50	727.83	0	0.51	2713.15	2734.94	99.73	99.67
45%	178	182	705.49	709.84	0	0.13	2719.91	2728.17	99.67	99.72

Next, we consider the two firms in $\mathcal{K}$ are asymmetric with respect to (α_A, α_B), i.e., $\alpha_A \neq \alpha_B$. We fix $\alpha_B = 45\%$ and vary α_A from 10% to 40%; that is, Firm B is the strong firm and Firm A is the weak firm. For ease of exposition, we define *Symmetry Level* (denoted by *SL*) as α_A/α_B, measuring the similarity between Firm A and Firm B. We further consider the factors that reflect the strong firm's ability to attract more loyal consumers. Specifically, the strong firm may use more advanced vehicles and provides better services during operations, leading to a higher unit allocation cost and higher penalty, relocation, and idle costs, respectively. Thus, we set higher cost parameters for the strong firm in the following experiments by multiplying the strong firm's cost parameters (i.e., unit allocation, penalty, relocation, and idle costs) by a factor of $u \geq 1$. Detailed managerial insights are provided in the following sections.

4.4.1 Vehicle Transfers

First, we examine how the symmetry level SL and factor u affect the number of transferred vehicles. Figure 8 displays the number of transferred vehicles averaged over all the regions and periods with respect to different values of SL and u. No matter what u is, the number of transferred vehicles first increases and then decreases as SL increases. Specifically, when SL is small, i.e., α_A is small, the number of trip demands that Firm A should satisfy is small and can be mostly covered by her own allocated vehicles, leading to a small number of transferred vehicles. When SL grows, i.e., α_A increases, Firm A sees a greater demand to satisfy, leading to an increased need for available vehicles to satisfy demands and hence an increased number of transferred vehicles. When SL further grows, i.e., α_A becomes large, the number of disloyal consumers in the market shrinks. The weak firm, i.e., Firm A, becomes more comparable to her opponent and tends to build capacity to serve her loyal consumers. Meanwhile, the strong firm is reluctant to build a high capacity because the shrinkage of disloyal consumers in the market limits her benefits. It follows that both firms become less likely to share their vehicles.



(a) $SL = 0.55$ **(b)** Various SL and u **(c)** $u = 1.75$

Figure 8. Number of Transferred Vehicles

Next, we examine the *temporal* and *spatial* features of transferred vehicles in detail. We consider $u = 1.25$ and $\alpha_A = 0.25$ and show the system's performance in Figures 9 and 10. Note that we choose $u = 1.25$ because shared micromobility firms do not have such large cost differences and choose $\alpha_A = 0.25$ because the number of transferred vehicles is significant (see Figure 8). Figure 9 shows the number of transferred vehicles averaged over all service regions during the entire operational horizon. The number of transferred vehicles is substantial during peak hours (i.e., $t \in [12, 18]_{\mathbb{Z}}$ or 8:00–9:00) and remains low during other periods, which is consistent with the trend of trip demands in Figure 4. A high trip demand indeed leads to a substantial need for available vehicles and an increased number of transferred vehicles, and vice versa.

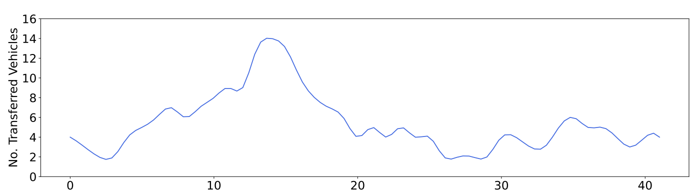
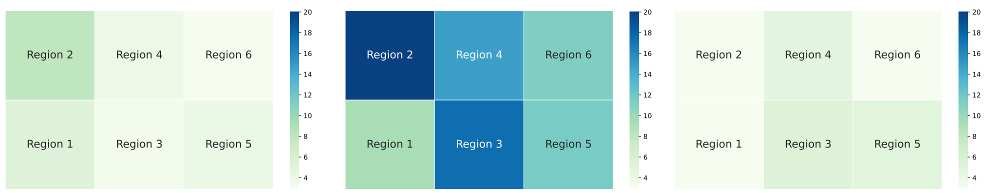


Figure 9. Vehicle Transfers

Figure 10 shows the number of transferred vehicles in each region in periods $t \in \{6, 18, 36\}$. Regions 2 and 3 are the regions where most vehicles are shared and transferred between the two firms. Note that these two regions also have the largest trip demands that originate from them (see Figure 5). Therefore, in periods and regions with high trip demands, we also see a large number of transferred vehicles. This indicates that capacity sharing can help firms to satisfy demands and potentially reduce traffic congestion on streets. We further examine the impacts of capacity sharing on firms' operations by comparing the results of Problem (9) and Problem (13) in the following three sections.



(a) $t = 6$ (7:00–7:10) **(b)** $t = 18$ (9:00–9:10) **(c)** $t = 36$ (12:00–12:10)

Figure 10. Spatial Features of Vehicle Transfers

4.4.2 Impact on Relocation

We examine the benefits of capacity sharing in reducing the number of relocated vehicles (i.e., *relocation reduction*). When capacity sharing is adopted, one might expect a firm to reduce her relocation because she can use vehicles from her opponent to satisfy demands instead of relocating her own vehicles. This expectation is confirmed by the results in Figure 11, showing that the total relocation reduction is positive and first increases then decreases with SL. This pattern with SL coincides with that in the number of transferred vehicles (see Figure 8). To obtain detailed insights into this pattern, we further investigate the *temporal* and *spatial* features of relocation reduction by fixing $u = 1.25$ and $\alpha_A = 0.25$.

Figure 12 shows the total number of relocated vehicles (i.e., *total relocation*) when capacity sharing is not considered (see the blue line) and the total relocation reduction (see the green/red bars) when capacity sharing is adopted in each period. The relocation reduction is positive in most periods and slightly negative in others. Meanwhile, the relocation reduction is limited during peak hours (i.e., $t \in [10, 20]_{\mathbb{Z}}$ or 7:40–9:20), but it is large right before and after the peak hours. To explain this phenomenon, we analyze the number of

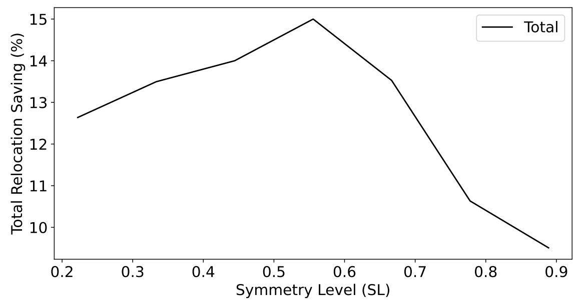


Figure 11. Impact of Capacity Sharing on Relocation Reduction

relocated vehicles that occur before, during, and after the peak hours when capacity sharing is not considered. (i) Before the peak hours, many vehicles are relocated, preparing for the upcoming high trip demands. After introducing capacity sharing, a firm can receive vehicles from her opponent to satisfy demands during peak hours (see many transferred vehicles during peak hours in Figure 9), by which the total relocation before the peak hours is reduced, i.e., the relocation reduction is significant. (ii) During the peak hours, the number of relocated vehicles is small because many vehicles are used to satisfy the high demands and few vehicles can be relocated. Meanwhile, since the demands decrease significantly after the peak hours, a large-scale relocation during peak hours is unnecessary. Due to the small number of relocated vehicles, the relocation reduction by capacity sharing is also limited. (iii) After the peak hours, the number of relocated vehicles becomes large. This happens because vehicles can be clustered in some regions after many consumer trips are completed and must be relocated to other regions for the upcoming demands there. After introducing capacity sharing, a firm can receive vehicles from her opponent and hence does not need to relocate as many vehicles as when capacity sharing is not adopted, leading to significant relocation reduction again.

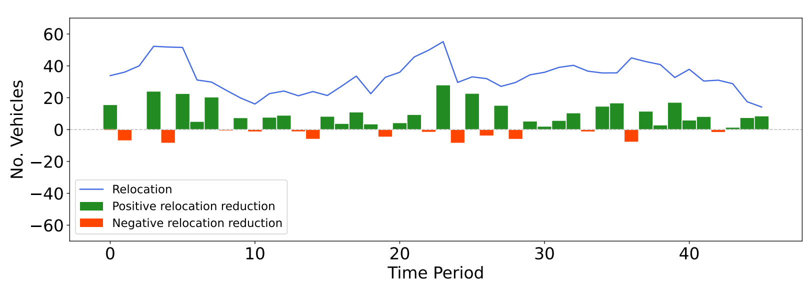
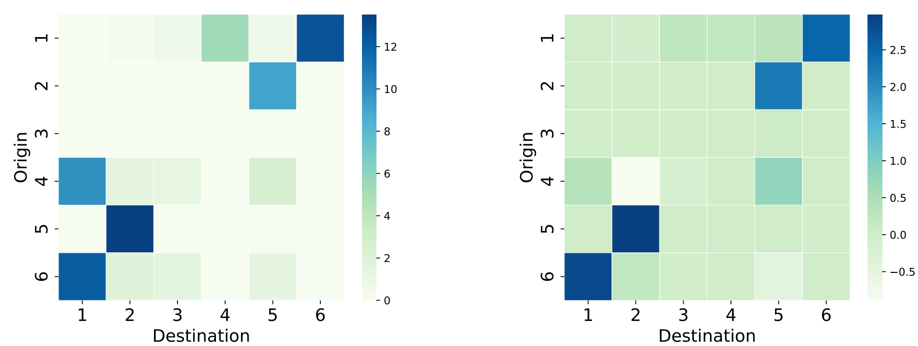


Figure 12. Impact of Capacity Sharing on Relocation Reduction in Each Period

Figure 13(a) displays the number of relocated vehicles in each region averaged over all the periods when capacity sharing is not adopted. We focus on the demand features in three clusters of regions (see Figures 5 and 6) and the corresponding relocation features. (i) In the first cluster with regions 2 and 3, which are the busiest pick-up regions, many consumers need to travel from there and hence each firm relocates many vehicles to this cluster to satisfy demands. (ii) In the second cluster with regions 4 and 5, which are the busiest destination regions, vehicles are cluttered there and hence each firm relocates many vehicles from this cluster to other regions. (iii) In the third cluster with regions 1 and 6, which are neither the busiest pick-up regions nor the busiest destination regions, each firm relocates vehicles both from and to this cluster. Figure 13(b) shows the relocation reduction in each region averaged over all the periods after introducing capacity sharing. Clearly, the relocation reduction mainly occurs in the regions where the total relocation is high, which confirms that *capacity sharing is a substitution for relocation*. We provide an example in Appendix K to demonstrate how capacity sharing can substitute for relocation.



(a) Spatial Features of Relocation

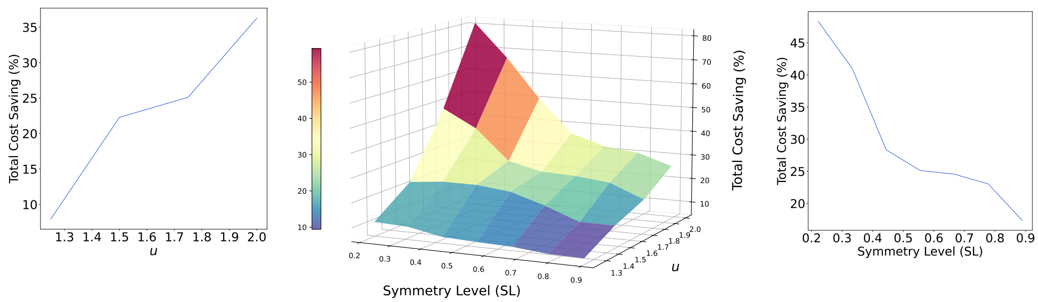
(b) Spatial Features of Relocation Reduction

Figure 13. Impact of Capacity Sharing on Relocation in Each Region

4.4.3 Impact on Total Costs

We further examine the benefits of capacity sharing in saving the firms' costs. The cost saving by capacity sharing is defined as the difference between the minimum total cost (i.e., the maximum total profit) without capacity sharing and that with capacity sharing. A positive cost saving means that capacity sharing improves the firms' profitability.

Figure 14 shows that the total cost saving over both firms is always positive, regardless of SL and u. This confirms the effectiveness of capacity sharing in introducing cost savings to the industry. Meanwhile, the cost-saving benefit shrinks as the two firms become more symmetric (i.e., SL becomes larger). Specifically, as α_A becomes closer to α_B, the weak firm (i.e., Firm A) tends to build her own capacity to serve her loyal and disloyal consumers. Meanwhile, the strong firm is also reluctant to build a high capacity because the shrinkage of disloyal consumers limits the benefits. Thus, firms are less likely to share their capacities when they become more symmetric, and the benefit of the capacity-sharing scheme is insubstantial compared to the case without capacity sharing. The above observations demonstrate two effects that the capacity-sharing scheme brings: (i) *the demand uncertainty reduction effect*, i.e., capacity sharing can align the excessive supply and demand; (ii) *the free-rider effect*, i.e., a weak firm with fewer loyal consumers can free-ride the strong opponent's capacity to serve consumers, especially disloyal consumers. More importantly, we should promote the capacity-sharing agreement when the two firms have different numbers of loyal consumers, specifically when the discrepancy is large.

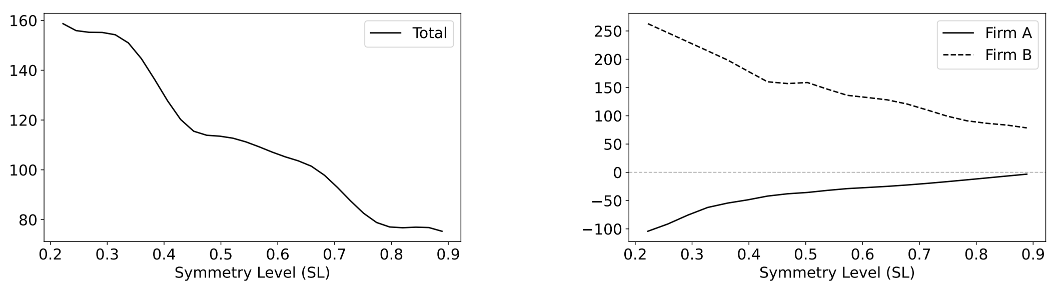


(a) $SL = 0.55$ **(b)** Various SL and u **(c)** $u = 1.75$

Figure 14. Impact of Capacity Sharing on Cost Saving

In addition, the cost saving increases with u. Note that capacity sharing can improve the two firms' operations, such as facilitating vehicle transfers (see Section 4.4.1) and reducing vehicle relocation (see Section 4.4.2). When the cost parameters of the strong firm increase, the cost that capacity sharing can reduce for her during the operation also increases. Thus, the cost-saving benefit is more significant as u increases, demonstrating that the unit cost of the strong firm affects the effectiveness of capacity sharing. We further investigate how capacity sharing affects each firm's total cost in detail. Specifically, we consider $u = 1.25$ and vary the value of SL.

Figure 15(a) continues to show that the total cost saving is positive and decreases with SL, whereas Figure 15(b) surprisingly shows that asymmetric firms cannot obtain better profitability simultaneously. Specifically, Firm A, the weak firm, has a negative cost saving, and Firm B, the strong firm, has a positive cost saving by capacity sharing. That is, capacity sharing benefits the strong firm and hurts the weak firm regarding cost reduction. Therefore, some interventions should be provided to incentivize both firms to reach a capacity-sharing agreement. For example, the strong firm that benefits may charge a lower transfer cost h_{T} from the weak firm. Moreover, the regulator may reduce some costs for the weak firm (i.e., Firm A), such as the plate registration cost (included by $\mathbf{c}^A$) and the idle cost (i.e., h_{I}^A), to incentivize the weak firm that suffers increased costs to reach the capacity-sharing agreement.



(a) Total Cost Saving

(b) Individual Cost Saving

Figure 15. Impact of Capacity Sharing on Cost Saving When $u = 1.25$

4.4.4 Impact of Budget

In the above studies, we set $\hat{x}^k = 1500$, a sufficiently large limit for firm $k \in \mathcal{K}$ to allocate vehicles. Here, we consider a budget limit that constrains firms from allocating vehicles. Specifically, we vary the values of $\hat{x}^A$ and $\hat{x}^B$ and examine how they affect the firms' demand loss and the effectiveness of capacity sharing in reducing demand loss. We let $u = 1.25$ and $(\alpha_A, \alpha_B) = (0.15, 0.45)$ and compare the results of Problems (9) and (13).

First, we fix $\hat{x}^A + \hat{x}^B = 360$ and vary the value of $\hat{x}^A$ from 160 to 60, by which $\hat{x}^B$ changes from 200 to 300. Figure 16 shows positive demand loss reduction, and such a reduction effect diminishes as $\hat{x}^B$ increases. Specifically, when $\hat{x}^B$ is small, Firm B, the strong firm, has large trip demands but is constrained from allocating enough vehicles to satisfy them, leading to significant demand losses. After capacity sharing is adopted, Firm

B can receive vehicles from her opponent to satisfy the demands, leading to significant demand loss reduction. When $\hat{x}^B$ increases, Firm B is allowed to allocate more vehicles to satisfy her trip demands, reducing her reliance on capacity sharing and alleviating the demand loss reduction effect.

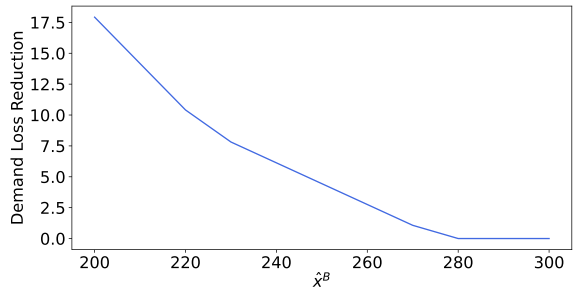
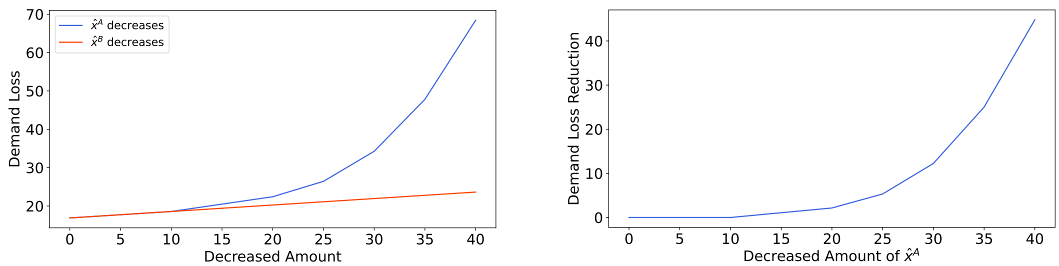


Figure 16. Demand Loss Reduction When $\hat{x}^A$ and $\hat{x}^B$ Change

Next, we examine the impact of allocation budgets by fixing $\hat{x}^k$ fixed and varying the value of $\hat{x}^{-k}$. We consider $(\hat{x}^A, \hat{x}^B) = (60, 300)$ and the following two cases: (i) fixing $\hat{x}^B = 300$ and decreasing the value of $\hat{x}^A$ from 60 to 20; (i) fixing $\hat{x}^A = 60$ and decreasing the value of $\hat{x}^B$ from 300 to 260. Figure 17(a) shows the total demand loss of the two firms for the above two cases when capacity sharing is not adopted. Specifically, decreasing $\hat{x}^A$ only (the weak firm) results in a greater demand loss than decreasing $\hat{x}^B$ only (the strong firm). When capacity sharing is further adopted, the total demand loss is reduced only for the first case. Figure 17(b) shows the corresponding demand loss reduction for the first case, where the reduction increases as $\hat{x}^A$ decreases. This implies that when the weak firm's allocation budget is severely limited, adopting capacity sharing can significantly reduce the entire market's demand loss. Therefore, we should promote the capacity-sharing agreement when the two firms have different allocation budgets, specifically when the discrepancy is large.

Remark 2 *In the above experiments, we assume that consumer demands do not change when capacity sharing is adopted. In Appendix M, we analyze the case where consumer demands are influenced by capacity sharing, and we further investigate its impact on two firms' performance in this case.*



(a) Demand Loss

(b) Demand Loss Reduction

Figure 17. Demand Loss When One Firm's Budget Changes

5 Conclusion

Unlike other shared mobility systems, a shared micromobility system often (i) bears the high investment cost of heavy assets (e.g., micromobility vehicles) while charging a low rental price, (ii) offers the convenience of vehicle pick-ups and drop-offs that may lead to a severe imbalance between supply and uncertain demand, and (iii) provides services not on an on-demand basis like ride sourcing, while the micromobility vehicles can be centrally managed and even shared with the competitors to serve demands. These features draw attention to the initial allocation and subsequent operation of shared micromobility vehicles. Currently, multiple shared micromobility firms operate in a city. The fierce competition among firms brings additional challenges to firms' decisions to allocate and relocate vehicles. Meanwhile, the city regulator (i) restricts the total number of vehicles in the market allocated to each service region to avoid traffic congestion on streets and (ii) expects to provide sustainable transportation services to as many city residents as possible. Such a restriction and expectation further challenge firms' vehicle allocation and relocation.

We consider two shared micromobility firms competing in the same service area with several service regions over an operational horizon. Each firm provides micromobility vehicles to satisfy the uncertain demands of two types of consumers: *loyal consumers* and *disloyal consumers*, capturing the demand heterogeneity in consumer loyalty. Each firm solves an integrated vehicle allocation and relocation problem, in which the total number of vehicles allocated by the two firms together in each service region is constrained by the city regulator, and provides a Nash equilibrium. Each firm's decision-making problem is formulated as a two-stage stochastic program on a spatial-temporal network, with the objective of maximizing her expected profit. In the capacity-development stage (1st stage), each firm decides on initial vehicle allocation for service regions. In the operation stage (2nd stage), the firm makes subsequent vehicle relocation decisions as recourse across service regions to match supply with demand after the demands are realized in each period. Specifically, consumer demands depend on capacity decisions (i.e., the number of allocated vehicles), capturing the demand features in the shared micromobility market.

We explore the optimality condition of each firm's decision-making and provide a tractable optimization formulation to obtain the Nash equilibrium by optimizing certain objectives (i.e., criteria for selecting an equilibrium) over the joint optimality conditions of both firms. In addition, to improve firms' operations with the limited number of allocated vehicles, we propose an innovative capacity-sharing agreement, under which a firm can share spare capacity for a fee with her opponent when both of them are competing in the same market. We prove that the capacity-sharing agreement between both firms benefits the entire system by reducing the expected total demand loss (see Proposition 1).

Based on real data collected from Citi Bike (2022), we perform numerical experiments to obtain managerial insights for the regulator and the firms with respect to two criteria for selecting an equilibrium: (i) minimizing the total demand loss of both firms and (ii) minimizing the total number of allocated vehicles. We discover the following insights that may provide valuable guidance for city regulations and for vehicle allocations and operations of shared micromobility firms in the competitive market.

(i) It is critical to set a proper initial vehicle allocation quota that approximates trip demands, by which each firm can achieve a high service level and profit simultaneously. A small quota may restrict firms from serving consumers, leading to a low service level, and a large quota may push firms to allocate many vehicles, leading to a high cost and traffic congestion (see Section 4.2).

(ii) The competition benefits the weak firm with a few loyal consumers because she can allocate more vehicles to earn a higher profit than before when no competition exists. Such a benefit is more significant when the number of disloyal consumers is large and a proper quota for initial vehicle allocation is enforced. Regarding vehicle allocation in each region, the strong firm dominates the busy regions with high consumer demands, while the weak firm focuses on non-busy regions. When the weak firm becomes more comparably competitive with the strong firm, the two firms' performance becomes similar (see Section 4.3).

(iii) With capacity sharing, one firm may act like a free rider that allocates fewer vehicles and asks for vehicles transferred from her opponent if needed when the two firms are symmetric. Meanwhile, many vehicles are shared in periods and regions with high demands. Capacity sharing can reduce the number of relocated vehicles by serving as a substitution for relocation and also improve the firms' profitability. However, the two firms cannot obtain better profitability simultaneously from the capacity-sharing agreement. Some interventions from the city regulator are needed to incentivize both firms to reach the agreement. In addition, capacity sharing can reduce the entire market's demand loss when the firms' allocation budgets are limited (see Section 4.4).

Although our numerical experiments are based on Citi Bike (2022), our proposed models and solution approaches are general enough for any typical shared micromobility market with two or more competing firms, thereby offering valuable insights for other shared micromobility systems with similar operational and competitive features. Our framework may also have potential applicability in car-sharing systems where the operator owns cars. In addition, we can apply our framework to a shared mobility system with electrical vehicles (EVs). Specifically, in addition to EV allocation and relocation, optimizing EV battery management during operations substantially increases the problem size and introduces significant computational challenges. Investigating the EV-sharing system within this framework and exploring methods to address the computational challenges may be of future interest. Note that the number of consumers firm $k \in \mathcal{K}$ attracts depends on the total number of allocated vehicles divided by the initial vehicle allocation quota (i.e., $\sum_{i \in \mathcal{M}} x_i^k / \sum_{i \in \mathcal{M}} \bar{x}_i$ in constraints (2)) because the quota approximates the size of trip demands, thereby manifesting the firm's ability in attracting consumers. We can also determine the number of consumers firm $k \in \mathcal{K}$ attracts based on her available vehicles. Numerical results indicate that this method performs similarly to our quota-based method but is more computationally difficult (see Appendix H). Furthermore, the loyal consumer percentages (α_A, α_B) are exogenously given because we focus on how the firms compete for disloyal consumers and further operate the system. It would be appealing to consider the decision to create loyal consumers. For instance, loyal consumers may be created by firms' advertising efforts and we can then endogenize the number of loyal consumers (Baye and Morgan, 2009) in our models. Finally, it would be interesting to study the impact of capacity sharing on the pattern of consumer demands in the shared micromobility system. We leave the above potential studies for future research.

Acknowledgements

The authors thank the Department Editor, Associate Editor, and two anonymous referees very much for their valuable comments and suggestions. This work was supported by the Research Grants Council of Hong Kong [Grants 15501920 and 15507718]. Ziliang Jin was supported in part by the Hong Kong PhD Fellowship Scheme when his work was done at the Hong Kong Polytechnic University and in part by the Start-up Research Fund of Southeast University [No. 4014002403].

Endnotes

1. We follow Jin et al. (2023) to estimate the vehicle allocation cost by dividing 150 dollars over 300 operational days and multiplying by 48/144 (i.e., scaling down the cost to our operational horizon), leading to $150/300 \times (48/144) \approx 0.16$.
2. We assume the revenue generated per trip in the area remains fixed, regardless of the trip's duration. This is because (i) Citi Bike charges a monthly fee of 17 from each consumer and allows this consumer to freely ride bikes for any trip in less than 45 minutes and (ii) all the trips in the area take less than 45 minutes. The revenue per trip is then estimated by assuming that each consumer has two rides per operational day, leading to $300/12 \times 2 = 50$ trips per month and $17/50 \approx 0.4$ per trip.
3. We test different values of h_{P}^k from 0.1 to 1, and the results are presented in Appendix L. The entire system's performance remains relatively consistent across different penalty values. Therefore, we set $h_{\mathrm{P}}^k = 0.1$ for simplicity.
4. Citi Bike's Bike Angels Rewards Program uses rewards to hire individual riders to relocate vehicles. Specifically, one point is rewarded after one relocation is completed, and every ten points can be exchanged for 1.2 dollars, leading to $1.2/10 = 0.12$ dollars per relocation.
5. The hourly car-parking rate in Manhattan is usually 4 dollars, leading to $4/6 \approx 0.6$ dollars per period (i.e., 10 min). A car-parking spot is assumed to accommodate 10 micromobility vehicles, leading to $0.6/10 = 0.06$ per vehicle per period.
6. The firm pays her opponent the transfer cost for receiving a vehicle to satisfy its demand. However, in cases where capacity sharing is not available, the firm may relocate a vehicle from another region to meet this demand. Thus, we assume that the transfer cost equals the relocation cost.

References

American Legal Publishing (2019). *Licensing and regulating shared mobility business*. Accessed: September 10, 2024.

Baye, M. R. and Morgan, J. (2009). Brand and price advertising in online markets. *Management Science*, 55(7):1139–1151.

Benjaafar, S., Wu, S., Liu, H., and Gunnarsson, E. B. (2022). Dimensioning on-demand vehicle sharing systems. *Management Science*, 68(2):1218–1232.

Cao, G., Jin, G. Z., Weng, X., and Zhou, L.-A. (2021). Market-expanding or market-stealing? Competition with network effects in bike-sharing. *The RAND Journal of Economics*, 52(4):778–814.

Chen, X., Fu, K., and Tang, Y. (2023). Impact of customer bounded rationality on on-demand service platforms. *IISE Transactions*, 55(10):1049–1061.

Chevalier-Roignant, B., Flath, C. M., Huchzermeier, A., and Trigeorgis, L. (2011). Strategic investment under uncertainty: A synthesis. *European Journal of Operational Research*, 215(3):639–650.

Citi Bike (2022). *Citi Bike Homepage*. Accessed: September 10, 2024.

City of Chicago (2022). *City of Chicago scooter sharing business rules*. Accessed: September 10, 2024.

Fishman, E., Washington, S., and Haworth, N. (2013). Bike share: A synthesis of the literature. *Transport Reviews*, 33(2):148–165.

Hargrove, D. and Ahn, R. (2022). *San Diego changes scooter and e-bike laws, caps amount allowed*. Accessed: September 10, 2024.

Hasija, S., Shen, Z.-J. M., and Teo, C.-P. (2020). Smart city operations: Modeling challenges and opportunities. *Manufacturing & Service Operations Management*, 22(1):203–213.

He, L., Mak, H.-Y., Rong, Y., and Shen, Z.-J. M. (2017). Service region design for urban electric vehicle sharing systems. *Manufacturing & Service Operations Management*, 19(2):309–327.

He, P., Zheng, F., Belavina, E., and Girotra, K. (2021a). Customer preference and station network in the london bike-share system. *Management Science*, 67(3):1392–1412.

He, Q.-C., Nie, T., Yang, Y., and Shen, Z.-J. (2021b). Beyond repositioning: Crowd-sourcing and geo-fencing for shared-mobility systems. *Production and Operations Management*, 30(10):3448–3466.

Jiang, Z., Lei, C., and Ouyang, Y. (2020). Optimal investment and management of shared bikes in a competitive market. *Transportation Research Part B: Methodological*, 135:143–155.

Jin, Z., Wang, Y., Lim, Y. F., Pan, K., and Shen, Z.-J. M. (2023). Vehicle rebalancing in a shared micromobility system with rider crowdsourcing. *Manufacturing & Service Operations Management*, 25(4):1394–1415.

Kabra, A., Belavina, E., and Girotra, K. (2020). Bike-share systems: Accessibility and availability. *Management Science*, 66(9):3803–3824.

Lu, M., Chen, Z., and Shen, S. (2017). Optimizing the profitability and quality of service in carshare systems under demand uncertainty. *Manufacturing & Service Operations Management*, 20(2):162–180.

Mahmoudi, M. and Zhou, X. (2016). Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state–space–time network representations. *Transportation Research Part B: Methodological*, 89:19–42.

Mak, H.-Y. (2022). Enabling smarter cities with operations management. *Manufacturing & Service Operations Management*, 24(1):24–39.

McKinsey (2019). *Micromobility's 15,000-mile checkup*. Accessed: September 10, 2024.

McKinsey (2021). *Why micromobility is here to stay*. Accessed: September 10, 2024.

Moug, K., Jia, H., and Shen, S. (2023). A shared-mobility-based framework for evacuation planning and operations under forecast uncertainty. *IISE Transactions*, 55(10):971–984.

Nair, R. and Miller-Hooks, E. (2011). Fleet management for vehicle sharing operations. *Transportation Science*, 45(4):524–540.

Nikzad, A. (2017). Thickness and competition in ride-sharing markets. *Available at SSRN 3065672*.

Oum, T. H., Park, J.-H., and Zhang, A. (1996). The effects of airline codesharing agreements on firm conduct and international air fares. *Journal of Transport Economics and Policy*, 30(2):187–202.

Qi, W., Li, L., Liu, S., and Shen, Z.-J. M. (2018). Shared mobility for last-mile delivery: Design, operational prescriptions, and environmental impact. *Manufacturing & Service Operations Management*, 20(4):737–751.

Qi, W. and Shen, Z.-J. M. (2019). A smart-city scope of operations management. *Production and Operations Management*, 28(2):393–406.

Reck, D. J., Martin, H., and Axhausen, K. W. (2022). Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D*, 102:103134.

Sago, D. V. (2020). *Bike sharing in China — From bicycle graveyards to a regulated industry*. Accessed: September 10, 2024.

Schellong, D., Sadek, P., Schaetzberger, C., and Barrack, T. (2019). *The promise and pitfalls of e-scooter sharing*. Accessed: September 10, 2024.

Schuijbroek, J., Hampshire, R. C., and Van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, 257(3):992–1004.

Shu, J., Chou, M. C., Liu, Q., Teo, C.-P., and Wang, I.-L. (2013). Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research*, 61(6):1346–1359.

Valdivia, A. G. (2020). *Barcelona's moped-sharing new licenses to challenge leading companies*. Accessed: September 10, 2024.

Van Mieghem, J. A. (2003). Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management*, 5(4):269–302.

Wang, J. and Lindsey, G. (2019). Do new bike share stations increase member use: A quasi-experimental study. *Transportation Research Part A: Policy and Practice*, 121:1–11.

Yu, M. and Shen, S. (2020). An integrated car-and-ride sharing system for mobilizing heterogeneous travelers with application in underserved communities. *IISE Transactions*, 52(2):151–165.

Zhang, C., Chen, J., and Raghunathan, S. (2022). Two-sided platform competition in a sharing economy. *Management Science*, 68(12):8909–8932.

Zhang, D., Yu, C., Desai, J., Lau, H., and Srivathsan, S. (2017). A time-space network flow approach to dynamic repositioning in bicycle sharing systems. *Transportation Research Part B: Methodological*, 103:188–207.

Supplemental Online Materials for "Integrated Vehicle Allocation and Relocation for Shared Micromobility under Competition and Demand Uncertainty" by Ziliang Jin, Kai Pan, Zuo-Jun Max Shen, and Wenxin Xu. *IISE Transactions*.

A Table of Notation

Table A1. Summary of Notation

Notation	Description
Sets:	
$\mathcal{K}$	set of shared micromobility firms $\mathcal{K} = \{A, B\}$
$\mathcal{M}$	set of service regions $\mathcal{M} = \{1, 2, \ldots, M\}$
$\mathcal{T}$	operational horizon $\mathcal{T} = \{0, 1, \ldots, T\}$
$\mathcal{G}$	time-space network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$
$\mathcal{N}$	set of nodes on the network $\mathcal{G}$
$\mathcal{E}$	set of directed arcs on the network $\mathcal{G}$
$\mathcal{E}^k_{\mathrm{R}}, \mathcal{E}^k_{\mathrm{I}}, \mathcal{E}^k_{\mathrm{L}}$	rental arcs, idle arcs, relocation arcs for each firm $k \in \mathcal{K}$
$\mathcal{E}^k_{\mathrm{T}}, \mathcal{E}^k_{\mathrm{N}_\delta}$	transfer arcs, return arcs for each firm $k \in \mathcal{K}$ with $\delta \in \{1, 2, \ldots, \Delta\}$
$\mathcal{A}$	$\mathcal{A} = \{\mathrm{R}, \mathrm{I}, \mathrm{L}, \mathrm{T}, \mathrm{N}_\delta, \ \forall \delta = 1, 2, \ldots, \Delta\}$
$\mathcal{Z}_{\mathrm{R}}$	$\mathcal{Z}_{\mathrm{R}} = \{\mathcal{M} \times \mathcal{M} \times \mathcal{T} \times \mathcal{T} : \ i \neq i', \ t' = t + \ell_{i,i'}\}$
$\mathcal{Z}_{\mathrm{L}}$	$\mathcal{Z}_{\mathrm{L}} = \{\mathcal{M} \times \mathcal{M} \times \mathcal{T} \times \mathcal{T} : \ i \neq i', \ t' = t + r_{i,i'}\}$
$\mathcal{Z}_{\mathrm{I}}$	$\mathcal{Z}_{\mathrm{I}} = \{\mathcal{M} \times \mathcal{T} : \ t \in \mathcal{T} \setminus \{T\}\}$
$\mathcal{S}$	set of scenarios of uncertain demands in all the service regions across all the periods
$\mathcal{W}_0$	feasible region of problems (9) and (10)
$\mathcal{W}_1$	feasible region of problems (13) and (14)
Parameters:	
$\ell_{i,i'}$	duration for a rental trip from region $i \in \mathcal{M}$ to region $i' \in \mathcal{M}$
$r_{i,i'}$	duration to relocate a vehicle from region $i \in \mathcal{M}$ to region $i' \in \mathcal{M}$
α_k	the percentage of consumers who are loyal to firm $k \in \mathcal{K}$
$\hat{x}^k$	upper bound for the number of vehicles allocated by firm $k \in \mathcal{K}$
$\bar{x}_i$	upper bound for the number of vehicles allocated to region $i \in \mathcal{M}$ by both firms in set $\mathcal{K}$
h^k_{R}	revenue of firm $k \in \mathcal{K}$ for serving a consumer
h^k_{P}	penalty cost of firm $k \in \mathcal{K}$ per consumer lost
h^k_{I}	idle cost of firm $k \in \mathcal{K}$ (e.g., the maintenance cost of idle vehicles) for each idling vehicle
h^k_{L}	cost for relocation a vehicle for firm $k \in \mathcal{K}$
h_{T}	cost of receiving a vehicle for firm $k \in \mathcal{K}$ from firm $-k$
c^k_i	cost incurred for allocating a vehicle for firm $k \in \mathcal{K}$ to region $i \in \mathcal{M}$
$D_{i,i',t,t'}$	the total market demand over any given time-space range $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$
Y^k_e	the number of loyal consumers of firm $k \in \mathcal{K}$ on each rental arc $e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}}$
$Y^0_{i,i',t,t'}$	the total number of disloyal consumers to both firms in set $\mathcal{K}$ over any given time-space range $(i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}$
θ_s	probability of each scenario $s \in \mathcal{S}$
ϵ	the minimum service level provided by each firm on all rental arcs over the whole operational horizon
Variables:	
x^k_i	number of vehicles allocated to region $i \in \mathcal{M}$ by firm $k \in \mathcal{K}$
w_e	realized flow on each arc $e \in \mathcal{E}$
$\overline{w}^k_e$	consumer demand (the number of consumers) of firm $k \in \mathcal{K}$ from region i in period t to region i' in period $t + \ell_{i,i'}$ for each rental arc $e = (n^k_{i,t}, n^k_{i',t+\ell_{i,i'}}) \in \mathcal{E}^k_{\mathrm{R}}$
$\tau_k, \beta_{k,i}, \kappa_{k,i}$	dual multipliers with respect to constraints (1)
$\rho_{k,i,i',t,t',s}, \gamma_{k,i,t,s}$	dual multipliers with respect to constraints (2)–(3)
$\eta^{\mathrm{R}}_{k,i,i',t,t',s}, \eta^{\mathrm{I}}_{k,i,t,s}$	dual multipliers with respect to constraints (4)–(5)
$v^{\mathrm{R}}_{k,i,i',t,t',s}, v^{\mathrm{L}}_{k,i,i',t,t',s}$	dual multipliers with respect to constraints (6)–(7)
$v^{\mathrm{I}}_{k,i,t,s}$	dual multipliers with respect to constraint (8)
$\mathbf{\Lambda}_0$	vector of all variables associated with both firms in $\mathcal{K}$ in problems (9) and (10)
$\mathbf{\Lambda}_1$	vector of all variables associated with both firms in $\mathcal{K}$ in problems (13) and (14)
Objective:	
Ψ^k	the profit of each firm $k \in \mathcal{K}$
$\Gamma^{\mathrm{DL}}_0 (\Gamma^{\mathrm{TA}}_0)$	the total demand loss (allocated vehicles) of both firms when the capacity-sharing agreement is not adopted
$\Gamma^{\mathrm{DL}}_1 (\Gamma^{\mathrm{TA}}_1)$	the total demand loss (allocated vehicles) of both firms when the capacity-sharing agreement is adopted

B KKT Conditions of Model $(\mathcal{D}^k(\mathbf{x}^{-k}))$, $\forall k \in \mathcal{K}$

First, we have the primal feasibility condition represented by the constraints (1)–(8). Second, to represent the dual feasibility condition, we define the following dual multipliers with respect to constraints (1)–(8):

$$\tau_k \geq 0, \; \forall k \in \mathcal{K}, \tag{15a}$$
$$\beta_{k,i} \geq 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \tag{15b}$$
$$\kappa_{k,i} \leq 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \tag{15c}$$
$$\rho_{k,i,i',t,t',s} \in \mathbb{R}, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \tag{15d}$$
$$\gamma_{k,i,0,s} \in \mathbb{R}, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; s \in \mathcal{S}, \tag{15e}$$
$$\gamma_{k,i,t,s} \in \mathbb{R}, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; t \in [1, T-1]_{\mathbb{Z}}, \; s \in \mathcal{S}, \tag{15f}$$
$$\gamma_{k,i,T,s} \in \mathbb{R}, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; s \in \mathcal{S}, \tag{15g}$$
$$\eta^{\mathrm{R}}_{k,i,i',t,t',s} \geq 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \tag{15h}$$
$$\eta^{\mathrm{I}}_{k,i,t,s} \geq 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; t \in [1, T-1]_{\mathbb{Z}}, \; s \in \mathcal{S}, \tag{15i}$$
$$v^{\mathrm{R}}_{k,i,i',t,t',s} \leq 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \tag{15j}$$
$$v^{\mathrm{L}}_{k,i,i',t,t',s} \leq 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{L}}, \; s \in \mathcal{S}, \tag{15k}$$
$$v^{\mathrm{I}}_{k,i,t,s} \leq 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; t \in [1, T-1]_{\mathbb{Z}}, \; s \in \mathcal{S}. \tag{15l}$$

Here, the multipliers defined in (15a)–(15i) correspond to constraints (1)–(7), respectively. The multipliers defined in (15j)–(15l), respectively, correspond to the following three parts of constraints (8): (i) $w_e \geq 0, \; \forall e \in \mathcal{E}^k_{\mathrm{R}}, \; k \in \mathcal{K}$, (ii) $w_e \geq 0, \; \forall e \in \mathcal{E}^k_{\mathrm{L}}, \; k \in \mathcal{K}$, and (iii) $w_e \geq 0, \; \forall e \in \mathcal{E}^k_{\mathrm{I}}, \; k \in \mathcal{K}$.

Third, we have the following constraints to represent the complementary slackness condition:

$$\left(\sum_{i \in \mathcal{M}} x^k_i - \hat{x}^k \right) \tau_k = 0, \; \forall k \in \mathcal{K}, \tag{16a}$$
$$\left(x^k_i + x^{-k}_i - \bar{x}_i \right) \beta_{k,i} = 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \tag{16b}$$
$$x^k_i \kappa_{k,i} = 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \tag{16c}$$
$$\left(w^{\mathrm{R}}_{k,i,i',t,t',s} - \overline{w}^{\mathrm{R}}_{k,i,i',t,t',s} \right) \eta^{\mathrm{R}}_{k,i,i',t,t',s} = 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \tag{16d}$$
$$\left(w^{\mathrm{I}}_{k,i,t,s} - x^k_i \right) \eta^{\mathrm{I}}_{k,i,t,s} = 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; t \in [1, T-1]_{\mathbb{Z}}, \; s \in \mathcal{S}, \tag{16e}$$
$$w^{\mathrm{R}}_{k,i,i',t,t',s} v^{\mathrm{R}}_{k,i,i',t,t',s} = 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \tag{16f}$$
$$w^{\mathrm{L}}_{k,i,i',t,t',s} v^{\mathrm{L}}_{k,i,i',t,t',s} = 0, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{L}}, \; s \in \mathcal{S}, \tag{16g}$$
$$w^{\mathrm{I}}_{k,i,t,s} v^{\mathrm{I}}_{k,i,t,s} = 0, \; \forall k \in \mathcal{K}, \; i \in \mathcal{M}, \; t \in [1, T-1]_{\mathbb{Z}}, \; s \in \mathcal{S}, \tag{16h}$$

where we rewrite

$$w^{\mathrm{R}}_{k,i,i',t,t',s} := w^s_e, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}, \; s \in \mathcal{S}, \; e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}},$$
$$w^{\mathrm{L}}_{k,i,i',t,t',s} := w^s_e, \; \forall k \in \mathcal{K}, \; (i,i',t,t') \in \mathcal{Z}_{\mathrm{L}}, \; s \in \mathcal{S}, \; e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{L}},$$
$$w^{\mathrm{I}}_{k,i,t,s} := w^s_e, \; \forall k \in \mathcal{K}, \; (i,t) \in \mathcal{Z}_{\mathrm{I}}, \; s \in \mathcal{S}, \; e = (n^k_{i,t}, n^k_{i,t+1}) \in \mathcal{E}^k_{\mathrm{I}}.$$

Fourth, we derive the Lagrangian stationarity conditions. Note that for each $k \in \mathcal{K}$, the Lagrangian function of model ($\mathcal{D}^k(\mathbf{x}^{-k})$) can be written as:

$$
\begin{aligned}
\mathcal{L}^k = & \sum_{i\in\mathcal{M}} \left(c_i^k + \tau_k + \beta_{k,i} - \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{R}} \sum_{s\in\mathcal{S}} \frac{Y^0_{i,i',t,t',s}}{\sum_{i\in\mathcal{M}} \bar{x}_i} \rho_{k,i,i',t,t',s} + \sum_{s\in\mathcal{S}} \left(\gamma_{k,i,T,s} - \gamma_{k,i,0,s} - \sum_{t=0}^{T-1} \eta^\mathrm{I}_{k,i,t,s} \right) + \kappa_{k,i} \right) x_i^k \\
& + \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{R}} \sum_{s\in\mathcal{S}} \left(\theta_s h_\mathrm{P}^k + \rho_{k,i,i',t,t',s} - \eta^\mathrm{R}_{k,i,i',t,t',s} \right) \overline{w}_{k,i,i',t,t',s} \\
& + \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{R}} \sum_{s\in\mathcal{S}} \left(-\theta_s h_\mathrm{R}^k - \theta_s h_\mathrm{P}^k + \gamma_{k,i,t,s} - \gamma_{k,i',t',s} + \eta^\mathrm{R}_{k,i,i',t,t',s} + v^\mathrm{R}_{k,i,i',t,t',s} \right) w^\mathrm{R}_{k,i,i',t,t',s} \\
& + \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{L}} \sum_{s\in\mathcal{S}} \left(\theta_s h_\mathrm{L}^k + \gamma_{k,i,t,s} - \gamma_{k,i',t',s} + v^\mathrm{L}_{k,i,i',t,t',s} \right) w^\mathrm{L}_{k,i,i',t,t',s} \\
& + \sum_{(i,t)\in\mathcal{Z}_\mathrm{I}} \sum_{s\in\mathcal{S}} \left(\theta_s h_\mathrm{I}^k + \gamma_{k,i,t,s} - \gamma_{k,i,t+1,s} + \eta^\mathrm{I}_{k,i,t,s} + v^\mathrm{I}_{k,i,t,s} \right) w^\mathrm{I}_{k,i,t,s} \\
& - \hat{x}^k \tau_k + \sum_{i\in\mathcal{M}} \left(x_i^{-k} - \bar{x}_i \right) \beta_{k,i} - \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{R}} \sum_{s\in\mathcal{S}} Y^k_{i,i',t,t',s} \rho_{k,i,i',t,t',s}. \qquad (17)
\end{aligned}
$$

Thus, the stationarity conditions can be described by the following constraints:

$$
-c_i^k = \tau_k + \beta_{k,i} - \sum_{(i,i',t,t')\in\mathcal{Z}_\mathrm{R}} \sum_{s\in\mathcal{S}} \frac{Y^0_{i,i',t,t',s}}{\sum_{i\in\mathcal{M}} \bar{x}_i} \rho_{k,i,i',t,t',s} + \sum_{s\in\mathcal{S}} \left(\gamma_{k,i,T,s} - \gamma_{k,i,0,s} - \sum_{t=0}^{T-1} \eta^\mathrm{I}_{k,i,t,s} \right) + \kappa_{k,i}, \quad \forall\, k\in\mathcal{K},\ i\in\mathcal{M}, \qquad (18a)
$$

$$
-\theta_s h_\mathrm{P}^k = \rho_{k,i,i',t,t',s} - \eta^\mathrm{R}_{k,i,i',t,t',s},\ \forall\, k\in\mathcal{K},\ (i,i',t,t')\in\mathcal{Z}_\mathrm{R},\ s\in\mathcal{S}, \qquad (18b)
$$

$$
\theta_s h_\mathrm{R}^k + \theta_s h_\mathrm{P}^k = \gamma_{k,i,t,s} - \gamma_{k,i',t',s} + \eta^\mathrm{R}_{k,i,i',t,t',s} + v^\mathrm{R}_{k,i,i',t,t',s},\ \forall\, k\in\mathcal{K},\ (i,i',t,t')\in\mathcal{Z}_\mathrm{R},\ s\in\mathcal{S}, \qquad (18c)
$$

$$
-\theta_s h_\mathrm{L}^k = \gamma_{k,i,t,s} - \gamma_{k,i',t',s} + v^\mathrm{L}_{k,i,i',t,t',s},\ \forall\, k\in\mathcal{K},\ (i,i',t,t')\in\mathcal{Z}_\mathrm{L},\ s\in\mathcal{S}, \qquad (18d)
$$

$$
-\theta_s h_\mathrm{I}^k = \gamma_{k,i,t,s} - \gamma_{k,i,t+1,s} + \eta^\mathrm{I}_{k,i,t,s} + v^\mathrm{I}_{k,i,t,s},\ \forall\, k\in\mathcal{K},\ (i,t)\in\mathcal{Z}_\mathrm{I},\ s\in\mathcal{S}. \qquad (18e)
$$

We use $\mathcal{W}_0$ to denote the feasible region defined by the above KKT conditions, including (1) for any $k\in\mathcal{K}$, (2)–(8) for any $s\in\mathcal{S}$, (15a)–(15l), (16a)–(16h), and (18a)–(18e). That is, $\mathcal{W}_0$ characterizes the set of possible equilibria of the Nash game between Firm A and Firm B.

Note that the complementary slackness conditions (16a)–(16h) are nonlinear. We can exactly linearize them for tractable computation. Specifically, we reformulate them into linear constraints by employing auxiliary binary variables. For example, for equation $x_i^k \kappa_{k,i} = 0$, given $k\in\mathcal{K},\ i\in\mathcal{M}$, we define a binary variable $z_{k,i}$ to indicate whether $\kappa_{k,i} = 0$ (i.e., $z_{k,i} = 1$) or $x_i^k = 0$ (i.e., $z_{k,i} = 0$). With an arbitrarily large positive number $G_{k,i}$, the equations $x_i^k \kappa_{k,i} = 0,\ \forall\, k\in\mathcal{K},\ i\in\mathcal{M}$ can be reformulated by the following linear constraints:

$$
-G_{k,i} z_{k,i} \le x_i^k \le G_{k,i} z_{k,i},\ \forall\, k\in\mathcal{K},\ i\in\mathcal{M}, \qquad (19)
$$

$$
-G_{k,i}(1-z_{k,i}) \le \kappa_{k,i} \le G_{k,i}(1-z_{k,i}),\ \forall\, k\in\mathcal{K},\ i\in\mathcal{M}, \qquad (20)
$$

$$
z_{k,i} \in \{0,1\},\ \forall\, k\in\mathcal{K},\ i\in\mathcal{M}. \qquad (21)
$$

Following the same way, we can also reformulate the remaining equations (16a)–(16h). We omit the details for simplicity.

C Vector Definition for Problem (9)

For Problem (9), we let $\mathbf{x} = (x_i^k,\ \forall\, i\in\mathcal{M}, k\in\mathcal{K})^\top$ denote the vector of both firms' initial vehicle allocation decisions in all regions, and $\mathbf{w} = (w_e^s,\ \forall\, e\in\cup_{k\in\mathcal{K}}\mathcal{E}^k,\ \overline{w}_e^s,\ \forall\, e\in\cup_{k\in\mathcal{K}}\mathcal{E}^k_\mathrm{R}, s\in\mathcal{S})^\top$

denote the vector of both firms' subsequent operational decisions on all types of arcs and in all scenarios. We further let $\boldsymbol{\tau} = (\tau_k,\ \forall\, k \in \mathcal{K})^\top$ denote the vector of dual multipliers defined in constraints (15a). For the other dual multipliers, we similarly let $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \boldsymbol{\rho},\ \boldsymbol{\gamma},\ \boldsymbol{\eta}^{\mathrm{R}},\ \boldsymbol{\eta}^{\mathrm{I}},\ \mathbf{v}^{\mathrm{R}},\ \mathbf{v}^{\mathrm{L}}$, and $\mathbf{v}^{\mathrm{I}}$ denote the vectors of dual multipliers defined in (15b)–(15l), respectively. In particular, the vectors $\boldsymbol{\beta},\ \boldsymbol{\kappa},\ \boldsymbol{\rho}$ denote the dual multipliers defined in (15b)–(15d), respectively. The vector $\boldsymbol{\gamma}$ denotes the dual multipliers defined in (15e)–(15g). The vectors $\boldsymbol{\eta}^{\mathrm{R}},\ \boldsymbol{\eta}^{\mathrm{I}},\ \mathbf{v}^{\mathrm{R}},\ \mathbf{v}^{\mathrm{L}}$, and $\mathbf{v}^{\mathrm{I}}$ denote the dual multipliers defined in (15h)–(15l), respectively. In addition, we let $\mathbf{y} = (\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\rho}, \boldsymbol{\eta}^{\mathrm{R}}, \boldsymbol{\eta}^{\mathrm{I}}, \mathbf{v}^{\mathrm{R}}, \mathbf{v}^{\mathrm{L}}, \mathbf{v}^{\mathrm{I}})^\top$ and $\boldsymbol{\Lambda}_0 = (\mathbf{x}, \mathbf{w}, \mathbf{y}, \boldsymbol{\gamma})^\top$. That is, we use $\boldsymbol{\Lambda}_0$ to denote a solution of Problem (9).

D KKT Conditions of Model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}_{\mathrm{C}}^{-k})$), $\forall\, k \in \mathcal{K}$

First, we have the primal feasibility condition represented by the constraints (1), (2)–(4) and (6)–(8) for any $s \in \mathcal{S}$, and (11)–(12). Second, to represent the dual feasibility condition, we continue to use the dual multipliers defined in (15a)–(15l), where $\gamma_{k,i,T,s} \in \mathbb{R},\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ s \in \mathcal{S}$ in constraints (15g) now correspond to constraints (11) that are updated from (5). In addition, we introduce the following new dual multipliers:

$$\zeta_{k,t,s} \in \mathbb{R},\ \forall\, k \in \mathcal{K},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{22}$$

$$v^{\mathrm{T}}_{k,i,t,s} \leq 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{23}$$

$$v^{\mathrm{E}}_{k,i,t,\delta,s} \leq 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{24}$$

where the multipliers in (22) correspond to the new constraints (12), the multipliers in (23) correspond to a new part $w^s_e \geq 0,\ \forall\, e \in \mathcal{E}^k_{\mathrm{T}},\ s \in \mathcal{S}$ in the updated constraints (8) for any $s \in \mathcal{S}$, and the multipliers in (24) correspond to a new part $w^s_e \geq 0,\ \forall\, e \in \mathcal{E}^k_{\mathrm{N}_\delta},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ t \in [1, T-1]_{\mathbb{Z}},\ k \in \mathcal{K},\ s \in \mathcal{S}$ in the updated constraints (8) for any $s \in \mathcal{S}$.

Third, to represent the complementary slackness condition, we continue to use constraints (16a)–(16h) and further introduce the following two constraints:

$$w^{\mathrm{T}}_{k,i,t,s} v^{\mathrm{T}}_{k,i,t,s} = 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{25}$$

$$w^{\mathrm{E}}_{k,i,t,\delta,s} v^{\mathrm{E}}_{k,i,t,\delta,s} = 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{26}$$

where

$$w^{\mathrm{T}}_{k,i,t,s} := w^s_e,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S},\ e = (n^k_{i,t}, n^{-k}_{i,t}) \in \mathcal{E}^k_{\mathrm{T}},$$

$$w^{\mathrm{E}}_{k,i,t,\delta,s} := w^s_e,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S},\ e = (n^k_{i,t}, n^{-k}_{i,t}) \in \mathcal{E}^k_{\mathrm{N}_\delta}.$$

Note that constraints (25)–(26) can also be reformulated as linear constraints following the same way in Appendix B.

Fourth, for the stationarity conditions, we continue to use constraints (18a)–(18e) and further introduce the following new constraints:

$$\theta_s h_{\mathrm{T}} = \gamma_{k,i,t,s} + v^{\mathrm{T}}_{k,i,t,s},\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{27}$$

$$0 = \gamma_{k,i,t+\delta,s} - \zeta_{k,t,s} + v^{\mathrm{E}}_{k,i,t,\delta,s},\quad \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S}. \tag{28}$$

In summary, the KKT conditions of the model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}_{\mathrm{C}}^{-k})$) for any $k \in \mathcal{K}$ are represented by constraints (1), (2)–(4) and (6)–(8) for any $s \in \mathcal{S}$, (15a)–(15l), (16a)–(16h), (18a)–(18e), (11)–(12), and (22)–(28). We use $\mathcal{W}_1$ to denote the feasible region defined by these constraints.

E Vector Definition for Problem (13)

For Problem (13), we adopt the vector definitions in C and make the following additional definitions. We let $\mathbf{w}^{\mathrm{T}} = (w_e^s,\ \forall\, e \in \cup_{k\in\mathcal{K}}\mathcal{E}_{\mathrm{T}}^k, s \in \mathcal{S})^\top$ denote the vector of both firms' operational decisions on all transfer arcs and in all scenarios, and $\mathbf{w}^{\mathrm{E}} = (w_e^s,\ \forall\, e \in \cup_{k\in\mathcal{K},\delta\in\{1,\dots,\Delta\}}\mathcal{E}_{\mathrm{N}_\delta}^k, s \in \mathcal{S})^\top$ denote the vector of both firms' operational decisions on all return arcs and in all scenarios. We further let $\boldsymbol{\zeta} = (\zeta_{k,t,s},\ \forall\, k \in \mathcal{K},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S})^\top$ denote the vector of dual multipliers defined in constraints (22). For the other dual multipliers, we similarly let $\boldsymbol{v}^{\mathrm{T}}$ and $\boldsymbol{v}^{\mathrm{E}}$ denote the vectors of dual multipliers defined in (23) and (24), respectively. In addition, we let $\boldsymbol{\Lambda}_1 = (\mathbf{x}, \mathbf{w}, \mathbf{y}, \gamma, \mathbf{w}^{\mathrm{T}}, \mathbf{w}^{\mathrm{E}}, \boldsymbol{v}^{\mathrm{T}}, \boldsymbol{v}^{\mathrm{E}}, \boldsymbol{\zeta})^\top$. That is, we use $\boldsymbol{\Lambda}_1$ to denote a solution of Problem (13).

F Proof of Proposition 1

Proof. Consider an optimal solution $\boldsymbol{\Lambda}_0^* = (\mathbf{x}^*, \mathbf{w}^*, \mathbf{y}^*, \gamma^*)^\top$ of Problem (9) with the corresponding optimal value Γ_0^{DL}. Based on this solution, we would like to construct a feasible solution $\hat{\boldsymbol{\Lambda}}_1 = (\hat{\mathbf{x}}, \hat{\mathbf{w}}, \hat{\mathbf{y}}, \hat{\gamma}, \hat{\mathbf{w}}^{\mathrm{T}}, \hat{\mathbf{w}}^{\mathrm{E}}, \hat{\boldsymbol{v}}^{\mathrm{T}}, \hat{\boldsymbol{v}}^{\mathrm{E}}, \hat{\boldsymbol{\zeta}})^\top$ of Problem (13).

We let $\hat{\mathbf{x}} = \mathbf{x}^*,\ \hat{\mathbf{w}} = \mathbf{w}^*,\ \hat{\mathbf{y}} = \mathbf{y}^*$, and $\hat{\mathbf{w}}^{\mathrm{T}} = \hat{\mathbf{w}}^{\mathrm{E}} = \mathbf{0}$. We also let

$$\hat{\gamma}_{k,i,t,s} = \gamma^*_{k,i,t,s} + \lambda_{k,s},\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M}, t \in \mathcal{T},\ s \in \mathcal{S}, \tag{29}$$

where

$$\lambda_{k,s} = \max_{\forall\, i\in\mathcal{M}, t\in[1,T-1]_{\mathbb{Z}}} (\theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s})^+,\ \forall\, k \in \mathcal{K},\ s \in \mathcal{S}, \tag{30}$$

and let

$$\hat{v}^{\mathrm{T}}_{k,i,t,s} = \theta_s h_{\mathrm{T}} - \hat{\gamma}_{k,i,t,s},\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M}, t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{31}$$

$$\hat{\zeta}_{k,t,s} = \min_{\forall\, i\in\mathcal{M}, \delta\in[1,\min\{\Delta, T-t\}]} \hat{\gamma}_{k,i,t+\delta,s},\ \forall\, k \in \mathcal{K},\ t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{32}$$

$$\hat{v}^{\mathrm{E}}_{k,i,t,\delta,s} = \hat{\zeta}_{k,t,s} - \hat{\gamma}_{k,i,t+\delta,s},\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S}. \tag{33}$$

It is easy to check that $\hat{\boldsymbol{\Lambda}}_1$ satisfies constraints (1) for any $k \in \mathcal{K}$, (2)–(4) and (6)–(8) for any $s \in \mathcal{S}$, (15a)–(15l), and (16a)–(16h). In addition, by (29), we have

$$\hat{\gamma}_{k,i,T,s} - \hat{\gamma}_{k,i,0,s} = \gamma^*_{k,i,T,s} + \lambda_{k,s} - \gamma^*_{k,i,0,s} - \lambda_{k,s} = \gamma^*_{k,i,T,s} - \gamma^*_{k,i,0,s}, \forall k \in \mathcal{K}, i \in \mathcal{M}, s \in \mathcal{S}, \tag{34a}$$

$$\hat{\gamma}_{k,i,t,s} - \hat{\gamma}_{k,i',t',s} = \gamma^*_{k,i,t,s} + \lambda_{k,s} - \gamma^*_{k,i',t',s} - \lambda_{k,s} = \gamma^*_{k,i,t,s} - \gamma^*_{k,i',t',s}, \forall k \in \mathcal{K}, (i, i', t, t') \in \mathcal{Z}_{\mathrm{R}}, s \in \mathcal{S}, \tag{34b}$$

$$\hat{\gamma}_{k,i,t,s} - \hat{\gamma}_{k,i,t+1,s} = \gamma^*_{k,i,t,s} + \lambda_{k,s} - \gamma^*_{k,i,t+1,s} - \lambda_{k,s} = \gamma^*_{k,i,t,s} - \gamma^*_{k,i,t+1,s}, \forall k \in \mathcal{K}, (i, t) \in \mathcal{Z}_{\mathrm{I}}, s \in \mathcal{S}. \tag{34c}$$

With (34a)–(34c), we can clearly show that $\hat{\boldsymbol{\Lambda}}_1$ satisfies constraints (18a)–(18e). Moreover, as $\hat{\mathbf{w}}^{\mathrm{T}} = \hat{\mathbf{w}}^{\mathrm{E}} = \mathbf{0}$, clearly $\hat{\boldsymbol{\Lambda}}_1$ satisfies (11)–(12). It is trivial that $\hat{\boldsymbol{\Lambda}}_1$ satisfies constraints (22).

For any $k \in \mathcal{K},\ i \in \mathcal{M}, t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}$, we have

$$\begin{aligned}\hat{v}^{\mathrm{T}}_{k,i,t,s} &= \theta_s h_{\mathrm{T}} - \hat{\gamma}_{k,i,t,s} = \theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s} - \lambda_{k,s} \le \theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s} - (\theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s})^+ \\ &= \begin{cases} 0, & \text{if } \theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s} > 0 \\ \theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s}, & \text{if } \theta_s h_{\mathrm{T}} - \gamma^*_{k,i,t,s} \le 0 \end{cases},\end{aligned} \tag{35}$$

where the first equality holds by (31), the second equality holds by (29) and $[1, T-1]_{\mathbb{Z}} \subseteq \mathcal{T}$, and the first inequality holds by (30). Clearly, (35) shows that $\hat{v}^{\mathrm{T}}_{k,i,t,s} \le 0,\ \forall\, k \in \mathcal{K},\ i \in \mathcal{M}, t \in [1, T-1]_{\mathbb{Z}},\ s \in \mathcal{S}$. Thus, $\hat{\boldsymbol{\Lambda}}_1$ satisfies constraints (23).

For any $k \in \mathcal{K},\ i \in \mathcal{M},\ t \in [1, T-1]_{\mathbb{Z}},\ \delta \in [1, \min\{\Delta, T-t\}]_{\mathbb{Z}},\ s \in \mathcal{S}$, we have

$$\hat{\vartheta}^{\mathrm{E}}_{k,i,t,\delta,s} = \hat{\zeta}_{k,t,s} - \hat{\gamma}_{k,i,t+\delta,s} = \min_{\forall i \in \mathcal{M}, \delta \in [1, \min\{\Delta, T-t\}]} \hat{\gamma}_{k,i,t+\delta,s} - \hat{\gamma}_{k,i,t+\delta,s} \leq 0,$$

where the first equality holds by (33) and the second equality holds by (32). Thus, $\hat{\mathbf{\Lambda}}_1$ satisfies constraints (24).

Since $\hat{\mathbf{w}}^{\mathrm{T}} = \hat{\mathbf{w}}^{\mathrm{E}} = \mathbf{0}$, it is trivial that $\hat{\mathbf{\Lambda}}_1$ satisfies constraints (25)–(26). By (31) and (33), we have $\hat{\mathbf{\Lambda}}_1$ satisfies constraints (27)–(28).

In summary, we have shown that $\hat{\mathbf{\Lambda}}_1$ satisfies all the constraints of Problem (13), including (1) for any $k \in \mathcal{K}$, (2)–(4) and (6)–(8) for any $s \in \mathcal{S}$, (15a)–(15l), (16a)–(16h), (18a)–(18e), (11)–(12), and (22)–(28). That is, $\hat{\mathbf{\Lambda}}_1$ is feasible to Problem (13), and the corresponding objective value equals

$$\sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_s \sum_{e \in \mathcal{E}^k_{\mathrm{R}}} \left(\hat{\overline{w}}^s_e - \hat{w}^s_e \right) = \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_s \sum_{e \in \mathcal{E}^k_{\mathrm{R}}} \left(\overline{w}^{s^*}_e - w^{s^*}_e \right) = \Gamma^{\mathrm{DL}}_0.$$

Given that Problem (13) has a feasible solution $\hat{\mathbf{\Lambda}}_1$, we have that the feasible region $\mathcal{W}_1$ is non-empty. Meanwhile, $\mathcal{W}_1$ is clearly bounded. It follows that Problem (13) has an optimal solution with the corresponding optimal value Γ^{DL}_1. Because Problem (13) is a minimization problem, the optimal value Γ^{DL}_1 is no larger than the objective value Γ^{DL}_0 with respect to the feasible solution $\hat{\mathbf{\Lambda}}_1$, i.e., $\Gamma^{\mathrm{DL}}_1 \leq \Gamma^{\mathrm{DL}}_0$. □

G Models Considering Loyal and Disloyal Consumers Separately

Our proposed models, such as model (13), consider loyal and disloyal consumers together during operations. Therefore, they minimize demand loss over all consumers, encompassing both loyal and disloyal ones, rather than ensuring satisfaction exclusively for either loyal or disloyal ones during operations. We can construct two models to separately address the needs of these two types of consumers. In particular, one model considers the loyal consumers only, while the other considers the disloyal consumers only.

First, in the model for loyal consumers, we introduce the following constraints to replace constraints (2) in model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}_{\mathrm{C}})$), thereby establishing the new demand constraints:

$$\overline{w}^s_e = Y^k_{i,i',t,t',s},\ \forall\, e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}},\ k \in \mathcal{K},\ s \in \mathcal{S}. \tag{36}$$

We then use $\mathcal{W}_{\mathrm{L}}$ to denote the feasible region defined by the KKT conditions of model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}_{\mathrm{C}})$), where constraints (2) are replaced with (36). Consequently, we formulate the following model to obtain the Nash equilibrium of two firms when they only consider loyal consumers:

$$\min_{\mathbf{\Lambda} \in \mathcal{W}_{\mathrm{L}}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_s \sum_{e \in \mathcal{E}^k_{\mathrm{R}}} \left(\overline{w}^s_e - w^s_e \right). \tag{37}$$

Meanwhile, through revising constraints (6) in $\mathcal{W}_{\mathrm{L}}$ to $w_e = \overline{w}_e$ for any $e \in \mathcal{E}^k_{\mathrm{R}}$ and $k \in \mathcal{K}$, model (37) ensure that all loyal consumers are satisfied.

Second, in the model for disloyal consumers, we introduce the following constraints to replace constraints (2) in model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}_{\mathrm{C}})$), thereby establishing the new demand constraints:

$$\overline{w}^s_e = Y^0_{i,i',t,t',s} \frac{\sum_{i \in \mathcal{M}} x^k_i}{\sum_{i \in \mathcal{M}} \bar{x}_i},\ \forall\, e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}},\ k \in \mathcal{K},\ s \in \mathcal{S}. \tag{38}$$

We then use $\mathcal{W}_{\mathrm{D}}$ to denote the feasible region defined by the KKT conditions of model ($\mathcal{C}^k(\mathbf{x}^{-k}, \mathbf{w}_{\mathrm{C}}^{-k})$), where constraints (2) are replaced with (38). Consequently, we formulate the following model to obtain the Nash equilibrium of two firms when they only consider disloyal consumers:

$$\min_{\mathbf{\Lambda}\in\mathcal{W}_{\mathrm{D}}} \sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}}\theta_s \sum_{e\in\mathcal{E}_{\mathrm{R}}^k} \left(\overline{w}_e^s - w_e^s\right). \tag{39}$$

Finally, we conduct extensive numerical experiments to test models (37) and (39) using the data introduced in Section 4.1 and present the results in Table G2. Table G2 compares the allocation, profit, and demand loss of model (13) (denoted by "Single Model") with the sum of results by models (37) and (39) without revising constraints (6) (denoted by "Separate Models"), and also with the sum of results by model (37) enforcing all loyal consumers are satisfied and model (39) (denoted by "Separate Satisfied Models"). We define

$$\mathrm{Gap}(\%) = \frac{\text{the result of Separate Models (Separate Satisfied Models)} - \text{the result of Single Model}}{\text{the result of Single Model}} \times 100\%.$$

Although the separate consideration of loyal and disloyal consumers ensures the satisfaction of loyal consumers, Table G2 suggests that considering loyal and disloyal consumers separately in two models results in lower profits and a greater demand loss compared to considering them together in a single model. Therefore, in our paper, we opt not to consider them separately.

Table G2. Performance of Different Models

	Allocation		**Profit ($)**		**Demand Loss**	
	Val	**Gap (%)**	**Val**	**Gap (%)**	**Val**	**Gap (%)**
Single Model	360	0	1408.69	0	16.88	0
Separate Models	360	0	1407.45	-0.09	16.88	0
Separate Satisfied Models	360	0	1405.72	-0.21	19.22	13.86

H Alternative Models Determining Consumers Based on Available Vehicles

We formulate an alternative equation to split the disloyal consumers $Y^0_{i,i',t,t',s}$ over any time-space range $(i,i',t,t') \in \mathcal{Z}_{\mathrm{R}}$ in any scenario $s \in \mathcal{S}$ to each firm $k \in \mathcal{K}$ based on their available vehicles in region i, time t, and scenario s, slightly different from equations (2).

Note that for each firm $k \in \mathcal{K}$, her available vehicles in region $i \in \mathcal{M}$, time $t \in \mathcal{T}$, and scenario $s \in \mathcal{S}$ are vehicles arriving there in t and s, i.e., $\sum_{e\in f^-(n^k_{i,t})} w_e^s$, the total vehicles on the rental, relocation, and idle arcs that terminate at $n^k_{i,t}$. We then provide the following alternative equation to determine the demand for each firm $k \in \mathcal{K}$:

$$\overline{w}_e^s = Y^0_{i,i',t,t',s}\frac{\sum_{e\in f^-(n^k_{i,t})} w_e^s}{\sum_{e\in f^-(n^k_{i,t})} w_e^s + \sum_{e\in f^-(n^{-k}_{i,t})} w_e^s} + Y^k_{i,i',t,t',s},\ \forall\, e = (n^k_{i,t}, n^k_{i',t'}) \in \mathcal{E}^k_{\mathrm{R}},\ s \in \mathcal{S}. \tag{40}$$

The ratio term $\sum_{e\in f^-(n^k_{i,t})} w_e^s / (\sum_{e\in f^-(n^k_{i,t})} w_e^s + \sum_{e\in f^-(n^{-k}_{i,t})} w_e^s)$ calculates the proportion of available vehicles of firm $k \in \mathcal{K}$ in region $i \in \mathcal{M}$, time $t \in \mathcal{T}$, and scenario $s \in \mathcal{S}$, relative to all available vehicles in the same region, time, and scenario. The larger it is, the more disloyal consumers firm k attracts. Constraints (40) are nonlinear due to the ratio term, thus requiring linearization. Before linearizing (40), we present the corresponding models using (40) for firms A and B, along with an algorithm to obtain their equilibrium.

Given firm $-k$'s decisions $\mathbf{x}^{-k}$ and $\mathbf{w}^{-k}$, we formulate the following alternative model for firm $k \in \mathcal{K}$, which considers the demand competition based on available vehicles.

$$\min \sum_{i\in\mathcal{M}} c_i^k x_i^k + \sum_{s\in\mathcal{S}} \theta_s \left(\sum_{e\in\mathcal{E}_{\mathrm{R}}^k} \left(h_{\mathrm{P}}^k(\overline{w}_e^s - w_e^s) - h_{\mathrm{R}}^k w_e^s \right) + \sum_{e\in\mathcal{E}_{\mathrm{I}}^k} h_{\mathrm{I}}^k w_e^s + \sum_{e\in\mathcal{E}_{\mathrm{L}}^k} h_{\mathrm{L}}^k w_e^s \right) \qquad (\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))$$
$$\text{s.t. } (1),\ (40);\ \ (3)-(8),\ \forall\, s \in \mathcal{S};\ \ w_e^s \in \mathbb{Z}_+,\ \forall\, e \in \mathcal{E}^k,\ s \in \mathcal{S}.$$

To obtain the equilibrium of firms A and B, we propose Algorithm 1 where models ($\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$), $k \in \mathcal{K}$ are solved iteratively, with subscript m denoting the iteration step.

Algorithm 1

Input: Initial solution $(\mathbf{x}_0^B, \mathbf{w}_0^B)$, iteration index $m = 0$.
Output: Equilibrium $(\mathbf{x}^{A^*}, \mathbf{w}^{A^*}, \mathbf{x}^{B^*}, \mathbf{w}^{B^*})$.
1: **do**
2: Solve Problem $(\mathcal{P}^A(\mathbf{x}_m^B, \mathbf{w}_m^B))$ to obtain an optimal solution $(\mathbf{x}_{m+1}^A, \mathbf{w}_{m+1}^A)$.
3: Solve Problem $(\mathcal{P}^B(\mathbf{x}_{m+1}^A, \mathbf{w}_{m+1}^A))$ to obtain an optimal solution $(\mathbf{x}_{m+1}^B, \mathbf{w}_{m+1}^B)$.
4: Set $m = m + 1$.
5: **while** $(\mathbf{x}_m^B, \mathbf{w}_m^B) \neq (\mathbf{x}_{m-1}^B, \mathbf{w}_{m-1}^B)$.
6: Set $(\mathbf{x}^{A^*}, \mathbf{w}^{A^*}, \mathbf{x}^{B^*}, \mathbf{w}^{B^*}) = (\mathbf{x}_m^A, \mathbf{w}_m^A, \mathbf{x}_m^B, \mathbf{w}_m^B)$.

In the following, we introduce how to linearize constraints (40) so that model ($\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$) of firm $k \in \mathcal{K}$ can be computationally tractable.

First, we introduce new variables $y_{i,t,s}^k, i \in \mathcal{M}, t \in \mathcal{T}, s \in \mathcal{S}$ to represent the ratio of firm k's available vehicles in region i, time t, and scenario s. Then we can reformulate constraints (40) as

$$\overline{w}_e^s = Y_{i,i',t,t',s}^0 y_{i,t,s}^k + Y_{i,i',t,t',s}^k,\ \forall\, e = (n_{i,t}^k, n_{i',t'}^k) \in \mathcal{E}_{\mathrm{R}}^k,\ s \in \mathcal{S}, \tag{41}$$
$$y_{i,t,s}^k = \frac{\sum_{e\in f^-(n_{i,t}^k)} w_e^s}{\sum_{e\in f^-(n_{i,t}^k)} w_e^s + \sum_{e\in f^-(n_{i,t}^{-k})} w_e^s},\ \forall\, i \in \mathcal{M},\ t \in \mathcal{T},\ s \in \mathcal{S}, \tag{42}$$

where constraints (42) are equivalent to

$$\sum_{e\in f^-(n_{i,t}^k)} w_e^s = \sum_{e\in f^-(n_{i,t}^k)} w_e^s y_{i,t,s}^k + \sum_{e\in f^-(n_{i,t}^{-k})} w_e^s y_{i,t,s}^k,\ \forall\, i \in \mathcal{M},\ t \in \mathcal{T},\ s \in \mathcal{S}. \tag{43}$$

Here, constraints (41) are linear, while constraints (43) are nonlinear due to bilinear terms $w_e^s y_{i,t,s}^k, e \in f^-(n_{i,t}^k), i \in \mathcal{M}, t \in \mathcal{T}, s \in \mathcal{S}$. Note that $w_e^s y_{i,t,s}^k, e \in f^-(n_{i,t}^{-k}), i \in \mathcal{M}, t \in \mathcal{T}, s \in \mathcal{S}$ are linear, since $w_e^s, e \in f^-(n_{i,t}^{-k}), i \in \mathcal{M}, t \in \mathcal{T}, s \in \mathcal{S}$ are given values from firm $-k$'s problem.

Second, before linearizing the bilinear terms $w_e^s y_{i,t,s}^k, e \in f^-(n_{i,t}^k), i \in \mathcal{M}, t \in \mathcal{T}, s \in \mathcal{S}$, we represent w_e^s using the binary expansion approach. Note that any finite integer can be represented in a unique binary format. For example, $w_e^s = 11$ can be converted to $w_e^s = 1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 1 \times 2^3$. In firm k's problem, it is obvious that w_e^s is bounded by $\sum_{i\in\mathcal{M}} \overline{x}_i$, i.e., the total capacity limit, for any $e \in f^-(n_{i,t}^k)$, $i \in \mathcal{M}$, $t \in \mathcal{T}$, and $s \in \mathcal{S}$. Without loss of generality, we assume that $\sum_{i\in\mathcal{M}} \overline{x}_i$ can be represented in binary format with $L+1$ digits. Then, for any $e \in f^-(n_{i,t}^k)$, $i \in \mathcal{M}$, $t \in \mathcal{T}$, and $s \in \mathcal{S}$, we introduce binary variables $z_{e,s}^l \in \{0,1\}, l = 0, \ldots, L$ and represent w_e^s by $w_e^s = \sum_{l=0}^L 2^l z_{e,s}^l$. Consequently, constraints (43) can be reformulated as

$$\sum_{e\in f^-(n_{i,t}^k)} \sum_{l=0}^{L} 2^l z_{e,s}^l = \sum_{e\in f^-(n_{i,t}^k)} \sum_{l=0}^{L} 2^l z_{e,s}^l y_{i,t,s}^k + \sum_{e\in f^-(n_{i,t}^{-k})} w_e^s y_{i,t,s}^k,\ \forall\, i \in \mathcal{M},\ t \in \mathcal{T},\ s \in \mathcal{S}, \tag{44}$$

which still have bilinear terms $z_{e,s}^l y_{i,t,s}^k, e \in f^-(n_{i,t}^k), i \in \mathcal{M}, t \in \mathcal{T}, l \in [0,L]_{\mathbb{Z}}, s \in \mathcal{S}$.

Third, we linearize these bilinear terms in constraints (44). By (42), we have $y_{i,t,s}^k \in [0,1]$ for any $i \in \mathcal{M}$, $t \in \mathcal{T}$, and $s \in \mathcal{S}$. Introducing variables $\sigma_{e,s}^l, e \in f^-(n_{i,t}^k), i \in \mathcal{M}, t \in \mathcal{T}, l \in [0,L]_{\mathbb{Z}}, s \in \mathcal{S}$, we can use the following McCormick inequalities to reformulate $\sigma_{e,s}^l = z_{e,s}^l y_{i,t,s}^k$:

$$0 \le \sigma_{e,s}^l \le z_{e,s}^l,\ \forall\, e \in f^-(n_{i,t}^k),\ i \in \mathcal{M},\ t \in \mathcal{T},\ l \in [0,L]_{\mathbb{Z}},\ s \in \mathcal{S}, \tag{45}$$

$$0 \le y_{i,t,s}^k - \sigma_{e,s}^l \le 1 - z_{e,s}^l,\ \forall\, e \in f^-(n_{i,t}^k),\ i \in \mathcal{M},\ t \in \mathcal{T},\ l \in [0,L]_{\mathbb{Z}},\ s \in \mathcal{S}. \tag{46}$$

When $z_{e,s}^l = 1$, constraints (45) hold and constraints (46) lead to $\sigma_{e,s}^l = y_{i,t,s}^k$. Conversely, when $z_{e,s}^l = 0$, constraints (45) lead to $\sigma_{e,s}^l = 0$ and constraints (46) hold.

Finally, by substituting $z_{e,s}^l y_{i,t,s}^k, e \in f^-(n_{i,t}^k), i \in \mathcal{M}, t \in \mathcal{T}, l \in [0,L]_{\mathbb{Z}}, s \in \mathcal{S}$ in constraints (44) with $\sigma_{e,s}^l$, in conjunction with constraints (45)–(46), we can reformulate constraints (44) as

$$\begin{aligned}&(45) - (46),\\ &\sum_{e \in f^-(n_{i,t}^k)} \sum_{l=0}^{L} 2^l z_{e,s}^l = \sum_{e \in f^-(n_{i,t}^k)} \sum_{l=0}^{L} 2^l \sigma_{e,s}^l + \sum_{e \in f^-(n_{i,t}^{-k})} w_e^s y_{i,t,s}^k,\ \forall\, i \in \mathcal{M},\ t \in \mathcal{T},\ s \in \mathcal{S}.\end{aligned} \tag{47}$$

It follows that constraints (40) in model ($\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$) for any $k \in \mathcal{K}$ can be equivalently reformulated as (41) and (45)–(47). Consequently, for any $k \in \mathcal{K}$, model ($\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$) can be reformulated as

$$\begin{aligned}\min\ & \sum_{i \in \mathcal{M}} c_i^k x_i^k + \sum_{s \in \mathcal{S}} \theta_s \left(\sum_{e \in \mathcal{E}_{\mathrm{R}}^k} \left(h_{\mathrm{P}}^k (\overline{w}_e^s - w_e^s) - h_{\mathrm{R}}^k w_e^s \right) + \sum_{e \in \mathcal{E}_{\mathrm{I}}^k} h_{\mathrm{I}}^k w_e^s + \sum_{e \in \mathcal{E}_{\mathrm{L}}^k} h_{\mathrm{L}}^k w_e^s \right) \qquad (\mathcal{P}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))\\ \text{s.t. }& (1);\ (3) - (8),\ \forall\, s \in \mathcal{S};\ (41),\ (45) - (47),\\ & w_e^s = \sum_{l=0}^{L} 2^l z_{e,s}^l,\ \forall\, e \in f^-(n_{i,t}^k),\ i \in \mathcal{M},\ t \in \mathcal{T},\ s \in \mathcal{S},\\ & w_e^s \in \mathbb{Z}_+,\ \forall\, e \in \mathcal{E}^k,\ s \in \mathcal{S};\ z_{e,s}^l \in \{0,1\},\ \forall\, e \in f^-(n_{i,t}^k),\ i \in \mathcal{M},\ t \in \mathcal{T},\ l \in [0,L]_{\mathbb{Z}},\ s \in \mathcal{S}.\end{aligned}$$

With Algorithm 1 and problems ($\mathcal{P}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$), $k \in \mathcal{K}$, we can then obtain the equilibrium of firms A and B when they split demands based on their available vehicles.

We further extend constraints (40) and model ($\mathcal{P}^k(\mathbf{x}^{-k}, \mathbf{w}^{-k})$) for each firm $k \in \mathcal{K}$ by considering capacity sharing between firms A and B. In addition to the rental, relocation, and idle vehicles arriving at region $i \in \mathcal{M}$ in time $t \in \mathcal{T}$ and scenario $s \in \mathcal{S}$, the available vehicles of firm k in region i, time t, and scenario s also include those returned from firm $-k$ under the capacity-sharing agreement. Note that we do not consider the vehicles transferred from firm $-k$ as available vehicles of firm $k \in \mathcal{K}$ because firms observe their demands generally before transferring vehicles. We let $\mathcal{R}(n_{i,t}^k)$ denote the set of return arcs that terminate at node $n_{i,t}^k \in \mathcal{N}$, and $f_{\mathrm{A}}^-(n_{i,t}^k)$ denote the set of rental, relocation, and idle arcs that terminate at node $n_{i,t}^k$. Then the equations determining the demand for each firm $k \in \mathcal{K}$ when capacity sharing is adopted are formulated as

$$\begin{aligned}\overline{w}_e^s = Y_{i,i',t,t',s}^0 \frac{\left(\sum_{e \in f_{\mathrm{A}}^-(n_{i,t}^k)} w_e^s + \sum_{e \in \mathcal{R}(n_{i,t}^k)} w_e^s\right)}{\left(\sum_{e \in f_{\mathrm{A}}^-(n_{i,t}^k)} w_e^s + \sum_{e \in \mathcal{R}(n_{i,t}^k)} w_e^s\right) + \left(\sum_{e \in f_{\mathrm{A}}^-(n_{i,t}^{-k})} w_e^s + \sum_{e \in \mathcal{R}(n_{i,t}^{-k})} w_e^s\right)} + Y_{i,i',t,t',s}^k,\\ \forall\, e = (n_{i,t}^k, n_{i',t'}^k) \in \mathcal{E}_{\mathrm{R}}^k,\ s \in \mathcal{S}.\end{aligned} \tag{48}$$

Consequently, with given firm $-k$'s decisions $\mathbf{x}^{-k}$ and $\mathbf{w}^{-k}$, we can formulate the following model for firm $k \in \mathcal{K}$ which considers the demand competition based on available vehicles, when capacity sharing is adopted.

$$\begin{aligned}
\min \quad & \sum_{i\in\mathcal{M}} c_i^k x_i^k + \sum_{s\in\mathcal{S}} \theta_s \Bigg(\sum_{e\in\mathcal{E}_\mathrm{R}^k} \left(h_\mathrm{P}^k(\overline{w}_e^s - w_e^s) - h_\mathrm{R}^k w_e^s \right) \qquad (\mathcal{Q}^k(\mathbf{x}^{-k},\mathbf{w}^{-k})) \\
& \qquad + \sum_{e\in\mathcal{E}_\mathrm{I}^k} h_\mathrm{I}^k w_e^s + \sum_{e\in\mathcal{E}_\mathrm{L}^k} h_\mathrm{L}^k w_e^s - \sum_{e\in\mathcal{E}_\mathrm{T}^k} h_\mathrm{T} w_e^s + \sum_{e\in\mathcal{E}_\mathrm{T}^{-k}} h_\mathrm{T} w_e^s \Bigg) \\
\text{s.t. } & (1),\ (48);\ \ (3)-(4),\ (6)-(8),\ (11)-(12),\ \forall\, s\in\mathcal{S};\ \ w_e^s \in \mathbb{Z}_+,\ \forall\, e\in\mathcal{E}^k,\ s\in\mathcal{S}.
\end{aligned}$$

Following the aforementioned linearization approach, we can reformulate model ($\mathcal{Q}^k(\mathbf{x}^{-k},\mathbf{w}^{-k})$) as

$$\begin{aligned}
\min \quad & \sum_{i\in\mathcal{M}} c_i^k x_i^k + \sum_{s\in\mathcal{S}} \theta_s \Bigg(\sum_{e\in\mathcal{E}_\mathrm{R}^k} \left(h_\mathrm{P}^k(\overline{w}_e^s - w_e^s) - h_\mathrm{R}^k w_e^s \right) \qquad (\mathcal{Q}_2^k(\mathbf{x}^{-k},\mathbf{w}^{-k})) \\
& \qquad + \sum_{e\in\mathcal{E}_\mathrm{I}^k} h_\mathrm{I}^k w_e^s + \sum_{e\in\mathcal{E}_\mathrm{L}^k} h_\mathrm{L}^k w_e^s - \sum_{e\in\mathcal{E}_\mathrm{T}^k} h_\mathrm{T} w_e^s + \sum_{e\in\mathcal{E}_\mathrm{T}^{-k}} h_\mathrm{T} w_e^s \Bigg) \\
\text{s.t. } & (1),\ (41);\ \ (3)-(4),\ (6)-(8),\ (11)-(12),\ \forall\, s\in\mathcal{S}, \\
& 0 \le \sigma_{e,s}^l \le z_{e,s}^l,\ \forall\, e\in f_\mathrm{A}^-(n_{i,t}^k)\cup\mathcal{R}(n_{i,t}^{-k}),\ i\in\mathcal{M},\ t\in\mathcal{T},\ l\in[0,L]_\mathbb{Z},\ s\in\mathcal{S}, \\
& 0 \le y_{i,t,s}^k - \sigma_{e,s}^l \le 1 - z_{e,s}^l,\ \forall\, e\in f_\mathrm{A}^-(n_{i,t}^k)\cup\mathcal{R}(n_{i,t}^{-k}),\ i\in\mathcal{M},\ t\in\mathcal{T},\ l\in[0,L]_\mathbb{Z},\ s\in\mathcal{S}, \\
& \sum_{e\in f_\mathrm{A}^-(n_{i,t}^k)} \sum_{l=0}^{L} 2^l z_{e,s}^l + \sum_{e\in\mathcal{R}(n_{i,t}^k)} w_e^s = \sum_{e\in f_\mathrm{A}^-(n_{i,t}^k)\cup\mathcal{R}(n_{i,t}^{-k})} \sum_{l=0}^{L} 2^l \sigma_{e,s}^l \\
& \qquad + \sum_{e\in f_\mathrm{A}^-(n_{i,t}^{-k})\cup\mathcal{R}(n_{i,t}^k)} w_e^s y_{i,t,s}^k,\ \forall\, i\in\mathcal{M},\ t\in\mathcal{T},\ s\in\mathcal{S}, \\
& w_e^s = \sum_{l=0}^{L} 2^l z_{e,s}^l,\ \forall\, e\in f_\mathrm{A}^-(n_{i,t}^k)\cup\mathcal{R}(n_{i,t}^{-k}),\ i\in\mathcal{M},\ t\in\mathcal{T},\ s\in\mathcal{S}, \\
& w_e^s \in \mathbb{Z}_+,\ \forall\, e\in\mathcal{E}^k,\ s\in\mathcal{S}, \\
& z_{e,s}^l \in \{0,1\},\ \forall\, e\in f_\mathrm{A}^-(n_{i,t}^k)\cup\mathcal{R}(n_{i,t}^{-k}),\ i\in\mathcal{M},\ t\in\mathcal{T},\ l\in[0,L]_\mathbb{Z},\ s\in\mathcal{S}.
\end{aligned}$$

Now, we conduct numerical experiments to test models ($\mathcal{P}_2^k(\mathbf{x}^{-k},\mathbf{w}^{-k})$) and ($\mathcal{Q}_2^k(\mathbf{x}^{-k},\mathbf{w}^{-k})$) for any $k\in\mathcal{K}$ by using Algorithm 1. These models are large-scale stochastic problems with numerous integer variables, presenting considerable computational complexity. Furthermore, Algorithm 1 requires numerous steps to achieve convergence. Consequently, we have to examine the performance of these models in a small-size case where the number of regions is 2, i.e., $|\mathcal{M}| = 2$, the number of periods is 3, i.e., $|\mathcal{T}| = 3$, and the number of scenarios is 3, i.e., $|\mathcal{S}| = 3$. The remaining parameters follow the settings presented in Section 4.1. We consider cases where two firms are either symmetric or asymmetric and capacity sharing is either adopted or not, leading to the following four cases:

Case (i): two symmetric firms with $(\alpha_A, \alpha_B) = (0.25, 0.25)$ when capacity sharing is not adopted.

Case (ii): two symmetric firms with $(\alpha_A, \alpha_B) = (0.25, 0.25)$ when capacity sharing is adopted.

Case (iii): two asymmetric firms with $(\alpha_A, \alpha_B) = (0.1, 0.45)$ when capacity sharing is not adopted.

Case (iv): two asymmetric firms with $(\alpha_A, \alpha_B) = (0.1, 0.45)$ when capacity sharing is adopted.

In cases (i) and (iii), we compare model (9) with $(\mathcal{P}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))$, $k \in \mathcal{K}$, while in cases (ii) and (iv), we compare model (13) with $(\mathcal{Q}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))$, $k \in \mathcal{K}$. For clarity in differentiation, we classify models that determine trip demands based on quota as "Quota Model" (e.g., models (9) and (13)), and models that determine demands based on available vehicles as "Availability Model" (e.g., models $(\mathcal{P}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))$ and $(\mathcal{Q}_2^k(\mathbf{x}^{-k}, \mathbf{w}^{-k}))$), $k \in \mathcal{K}$).

In Tables H3–H6, we compare the computational time of solving the quota model with the time used by Algorithm 1 to obtain the equilibrium of the availability models, as detailed in the column "Time (s)." In addition, we compare the equilibria obtained by the quota model and the availability model, focusing on the results for firm A, firm B, and both firms (see the columns "A," "B," and "Total"). This comparison encompasses: (i) amounts of allocated vehicles in regions 1 and 2 (see the columns "Region 1" and "Region 2"), (ii) expected demands across all regions and periods (see the column "Demand"), (iii) the expected operation profit, which equals the expected total profit minus the allocation cost (see the column "Operation Profit ($)"), and (iv) the expected number of transferred vehicles in all regions and periods (see the column "Transfer"). We calculate the gap between the results of these models by defining Gap $=$ (the result of Quota Model) $-$ (the result of Availability Model).

Table H3. Performance in Case (i)

	Time (s)	Region 1			Region 2			Demand			Operation Profit ($)		
		A	B	Total	A	B	Total	A	B	Total	A	B	Total
Quota Model	0.7	26	26	52	24	24	48	77	77	154	37.0	37.0	74.0
Availability Model	8.1	26	26	52	22	22	44	77	77	154	37.3	37.3	74.6
Gap	-7.4	0	0	0	2	2	4	0	0	0	-0.3	-0.3	-0.6

Table H4. Performance in Case (ii)

	Time (s)	Region 1			Region 2			Demand			Operation Profit ($)			Transfer		
		A	B	Total	A	B	Total	A	B	Total	A	B	Total	A	B	Total
Quota Model	2.3	25	27	52	21	27	48	74	80	154	34.9	39.1	74.0	0	0.5	0.5
Availability Model	61.7	23	29	52	19	24	43	71	83	154	34.5	40.3	74.8	0	0.7	0.7
Gap	-59.4	2	-2	0	2	3	5	3	-3	0	0.4	-1.2	-0.8	0	-0.2	-0.2

Table H5. Performance in Case (iii)

	Time (s)	Region 1			Region 2			Demand			Operation Profit ($)		
		A	B	Total	A	B	Total	A	B	Total	A	B	Total
Quota Model	0.7	12	40	52	12	36	48	32	122	154	14.4	52.1	66.5
Availability Model	82.1	12	39	51	11	33	44	33	121	154	16.1	53.9	70.0
Gap	-81.4	0	1	1	1	3	4	-1	1	0	-1.7	-1.8	-3.5

Table H6. Performance in Case (iv)

	Time (s)	Region 1			Region 2			Demand			Operation Profit ($)			Transfer		
		A	B	Total	A	B	Total	A	B	Total	A	B	Total	A	B	Total
Quota Model	4.3	9	43	52	10	38	48	29	125	154	9.6	57.5	67.1	0	3.1	3.1
Availability Model	697.1	7	43	50	7	33	40	27	127	154	12.1	58.3	70.4	0	4.0	4.0
Gap	-692.8	2	0	2	3	5	8	2	-2	0	-2.5	-0.8	-3.3	0	-0.9	-0.9

It is obvious that the availability model and the quota model generate equilibria that are very close. Specifically, the results obtained by these models show a high similarity across various aspects, including firms' allocations, demands, operation profit, and transfer amount when capacity sharing is adopted. We have also verified that the managerial insights obtained by the availability model and the quota model remain consistent. These findings imply that *the quota model is reliable*.

In addition, the quota model takes a significantly shorter computational time than the availability model, with a notable decrease of over 99.3% in case (iv). This substantial reduction in the computational time highlights *the quota model's practical applicability*. Moreover, the quota model splits demands based on the quota issued by the city regulator, thereby empowering the regulator to control vehicle allocations through quota adjustments. For example, the regulator can dynamically adjust the quota to control firms' allocations according to their performance in serving demands, which aligns with the practices in Chicago (City of Chicago, 2022). This mechanism allows the regulator to play a more active role in controlling vehicle allocations, thereby strengthening its regulatory capabilities and avoiding potential traffic congestion on the street.

I Summary of All Trip Demands

Figure 18 shows the summary of all the trip demands, where a subfigure in a cubic represents the trip demands from a region $i \in \mathcal{M}$ to another region $j \in \mathcal{M}$.

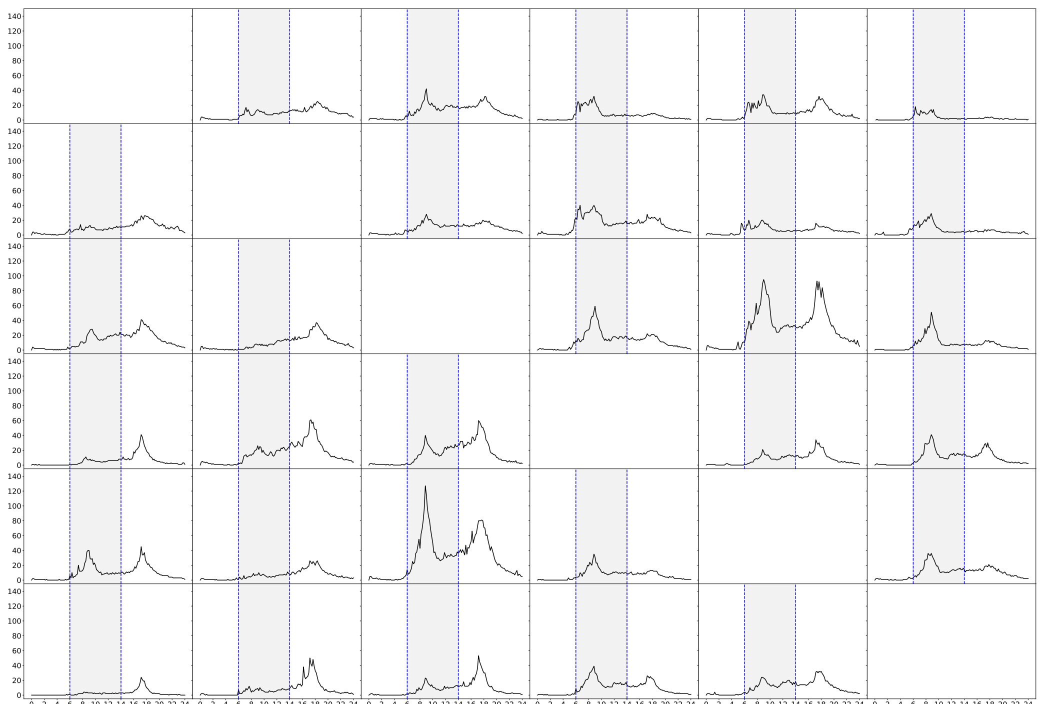


Figure 18. The Sketch of Demand

J Computational Performance

We consider Problem (9) and examine the impact of $|\mathcal{S}|$ on its computational performance, thereby demonstrating that setting $|\mathcal{S}| = 3$ is effective enough. Given the multivariate normal distribution $\mathcal{N}(\hat{\boldsymbol{D}}_t, \boldsymbol{\Sigma}_t)$ for any $t \in \mathcal{T}$, we let Ω represent the number of samples that we randomly generate from this distribution. Here, we consider $\Omega \in \{100 \times i, \forall i \in [1, 10]_{\mathbb{Z}}\}$. Given Ω randomly generated samples, we solve two instances of Problem (9): (i) $\mathcal{S}$ collects all the samples, with each sample representing a separate scenario, i.e., $|\mathcal{S}| = \Omega$; (ii) we use k-means clustering to cluster these Ω samples into three groups, with the center of each group representing a separate scenario, i.e., $|\mathcal{S}| = 3$. For each $i \in \mathcal{M}$ and $k \in \mathcal{K}$, we let $\hat{x}_i^{k*}$ (resp. x_i^{k*}) represent firm k's optimal initial allocation decision in region i after solving the first (resp. second) instance above. For each firm $k \in \mathcal{K}$, we use $(\sum_{i\in\mathcal{M}}(x_i^{k^*} - \hat{x}_i^{k^*})^2/|\mathcal{M}|)^{1/2}$ to measure the *relative error* of the solutions (i.e., solution difference) between the above two instances. Figure 19 shows the results.

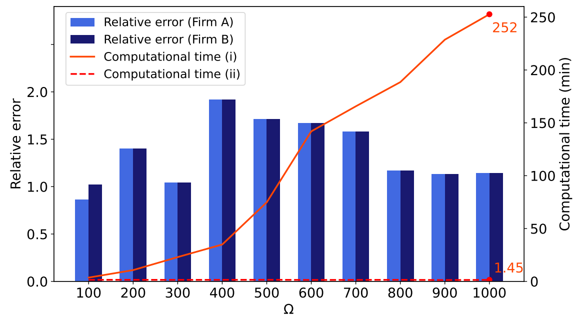


Figure 19. Relative Error and Computational Time

For each firm $k \in \mathcal{K}$, the *relative error* is around 1 and always less than 2, regardless of the value of Ω. Note that the optimal solutions of both the above two instances are to allocate 360 vehicles in all regions by the two firms, i.e., $\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{M}}\hat{x}_i^{k^*} = \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{M}}x_i^{k^*} = 360$, leading to 60 vehicles in each region on average. Thus, when the *relative error* is 1, it accounts for only 1.67% ($= 1/60 \times 100\%$) of all the allocated vehicles. Such a small error indicates that choosing three scenarios via clustering is effective and representative enough compared to choosing all the available samples. More importantly, Figure 19 shows that when $|\mathcal{S}|$ is larger, solving Problem (9) becomes more challenging. For instance, when $|\mathcal{S}| = 1,000$ in the first instance, the computational time (see the red solid line) is 252 minutes, while the computational time (see the red dashed line) is only 1.45 minutes when $|\mathcal{S}| = 3$ in the second instance. Thus, choosing three scenarios also helps us efficiently obtain the solutions, enabling practical applications in the industry. As a result, in the subsequent numerical experiments, we focus on using the three scenarios ($|\mathcal{S}| = 3$) we select in Section 4.1.

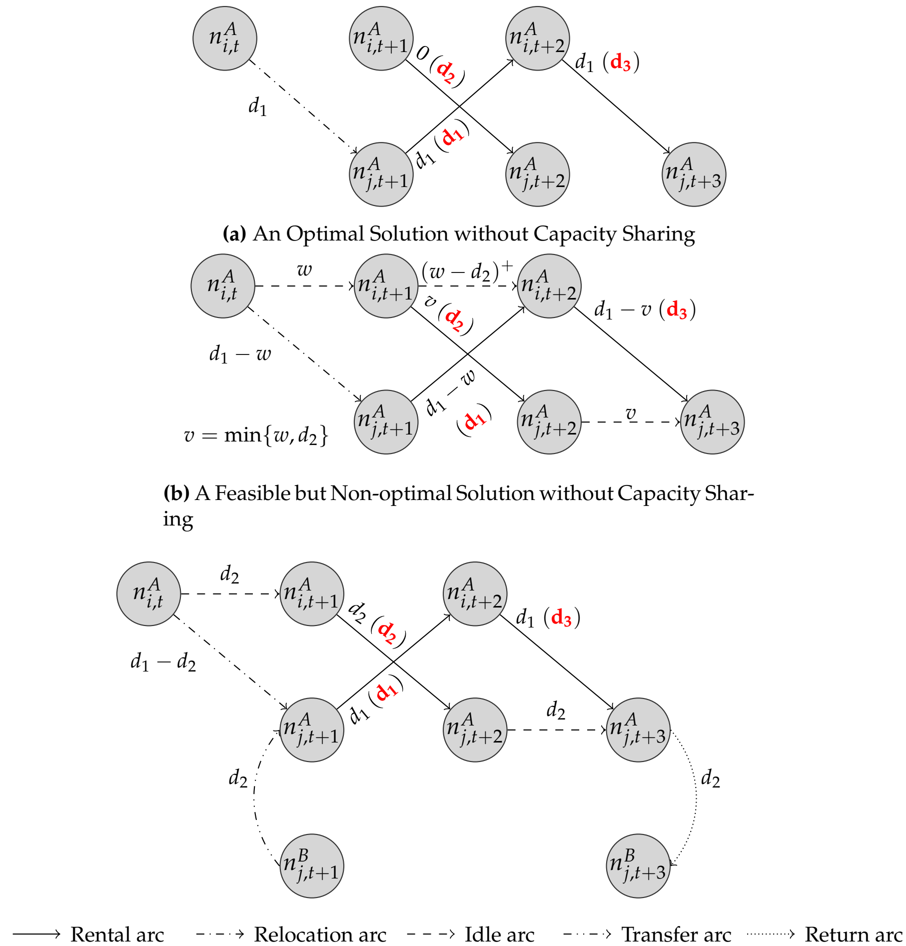
K Capacity Sharing as a Substitution for Relocation

Figure 20 provides an example to demonstrate how capacity sharing can substitute for relocation. We consider a simplified case where Firm A operates in regions i and j from periods t to $t+3$. The number of trip demands from region j to i in period $t+1$ is d_1, and the numbers of trip demands from region i to j in periods $t+1$ and $t+2$ are d_2 and d_3, respectively. We assume that $d_3 \geq d_1 \geq d_2$ and show the number of vehicles on each arc in Figure 20. For example, in Figure 20(a), the number of vehicles on the arc from $n_{i,t+1}^A$ to $n_{j,t+2}^A$ is 0, while the trip demand from $n_{i,t+1}^A$ to $n_{j,t+2}^A$ is d_2.

Figure 20(a) shows an optimal solution when capacity sharing is not adopted. Firm A relocates d_1 vehicles in advance from region i in period t to serve the d_1 consumers originating from region j in $t+1$. These vehicles are then reused to serve some of the d_3 consumers originating from region i in period $t+2$. Note that none of the d_2 consumers originating from region i in period $t+1$ is served. The total cost corresponding to this optimal solution $\text{cost}_\text{a} = h_\text{L}^A d_1 + h_\text{P}^A(d_2 + d_3 - d_1) - 2h_\text{R}^A d_1$. Figure 20(b) shows a solution with $d_1 - w$ vehicles relocated from region i in period t to region j when capacity sharing is not adopted, where $0 \leq w \leq d_1$, leaving w vehicles idle in region i and period t. The total cost corresponding to this solution

$$\begin{aligned}\text{cost}_\text{b} &= h_\text{L}^A(d_1 - w) + h_\text{P}^A(w + d_3 - d_1 + d_2) + 2h_\text{I}^A w - h_\text{R}^A(d_1 - w + d_1)\\ &= \text{cost}_\text{a} + \left(h_\text{R}^A + h_\text{P}^A + 2h_\text{I}^A - h_\text{L}^A\right) w.\end{aligned}$$

As $h_\text{R}^A + h_\text{P}^A + 2h_\text{I}^A - h_\text{L}^A > 0$ by the parameter setting in Section 4.1, we have $\text{cost}_\text{b} > \text{cost}_\text{a}$. This explains why Firm A relocates d_1 vehicles, instead of fewer ones, from region i in period t when



(a) An Optimal Solution without Capacity Sharing

(b) A Feasible but Non-optimal Solution without Capacity Sharing

(c) An Optimal Solution with Capacity Sharing

Figure 20. Capacity Sharing Reduces Relocation

capacity sharing is not adopted. In addition, following the solution in Figure 20(a), Firm A serves $2d_1$ consumers in total, but she serves only $2d_1 - w$ consumers following the solution in Figure 20(b). Thus, a good relocation decision helps the firms serve more consumers and maximize profit.

Figure 20(c) shows an optimal solution when capacity sharing is adopted. Firm A relocates $d_1 - d_2$ vehicles from region i in period t to region j, where $0 \leq d_2 \leq d_1$, and receives d_2 vehicles from her opponent (i.e., Firm B) to serve the d_1 consumers originating from j in $t+1$. Some of the d_3 consumers originating from i in $t+2$ are served, and the received d_2 vehicles are returned to Firm B in period $t+3$. The total cost corresponding to this solution

$$\begin{aligned} \text{cost}_\text{c} &= h_\text{L}^A(d_1 - d_2) + h_\text{P}^A(d_3 - d_1) + 2h_\text{I}^A d_2 + h_\text{T} d_2 - h_\text{R}^A(2d_1 + d_2) \\ &= \text{cost}_\text{a} - (h_\text{R}^A + h_\text{P}^A - 2h_\text{I}^A + h_\text{L}^A - h_\text{T})d_2. \end{aligned}$$

As $h_\text{R}^A + h_\text{P}^A - 2h_\text{I}^A + h_\text{L}^A - h_\text{T} > 0$ by the parameter setting in Section 4.1, we have $\text{cost}_\text{c} < \text{cost}_\text{a}$. Therefore, by adopting capacity sharing, a firm can reduce the number of relocated vehicles while satisfying more consumers (i.e., $2d_1 + d_2$) than that (i.e., $2d_1$) when capacity sharing is not adopted.

L Impact of Penalty and Demand Level

We conduct experiments by varying the penalty cost (i.e., h_{P}^k, $k \in \mathcal{K}$) and demand level to examine how different penalty costs affect firms' performance under different demand sizes. The demand level refers to the extent to which the trip demand is increased. For example, a level of 10 corresponds to multiplying the trip demand set in Section 4.1 by 10. With $(\alpha_A, \alpha_B) = (0.1, 0.45)$ and $u = 1.25$, we solve Problem (9) and Problem (13) and show the results in Table L7 and Table L8, respectively. We define

$$\text{Gap}(\%) = \frac{|\text{the result when penalty is } 0.1 - \text{the result when penalty is not } 0.1|}{\text{the result when penalty is } 0.1} \times 100\%,$$

with respect to each instance with a given demand level and penalty cost.

Table L7 shows that when the demand level is 1 and the penalty cost varies from 0.1 to 1, the Gap (%) of allocation, profit, demand loss, or relocation remains within 1%. This suggests that varying the penalty cost has a limited impact on firms' performance when the demand level is 1 and capacity sharing is not adopted. We observe very similar results in Table L8 where capacity sharing is adopted. The table shows that when the demand level is 1, the Gap (%) remains within 1.6% as the penalty cost varies.

When the demand level is 1, which aligns with the trip demand setting in Section 4.1, the two firms' performance is weakly affected by changes in penalty cost, no matter whether capacity sharing is adopted. Therefore, we set the penalty cost at 0.1 in numerical experiments for the sake of simplicity while ensuring the obtained results are promising and generalizable.

Table L7 and Table L8 also reveal that as the demand level increases, indicating trip demand rises, firms' performance is impacted by the penalty cost, specifically in terms of vehicle relocation and transfer. Therefore, it is crucial to carefully estimate the penalty cost in the case where the trip demand is very large.

Table L7. Performance Without Sharing

Level	Penalty	Allocation		Profit		Demand Loss		Relocation	
		Value	Gap (%)	Value ($)	Gap (%)	Value	Gap (%)	Value	Gap (%)
1	0.1	360	0	1250.32	0	16.88	0	75.00	0
	0.4	360	0	1246.86	0.28	16.88	0	75.29	0.39
	0.7	360	0	1243.20	0.57	16.88	0	75.31	0.42
	1	360	0	1241.51	0.71	16.88	0	75.44	0.58
4	0.1	1440	0	5000.32	0	67.53	0	305.38	0
	0.4	1440	0	4980.20	0.40	67.53	0	306.56	0.39
	0.7	1440	0	4950.80	0.99	67.53	0	307.06	0.55
	1	1440	0	4932.60	1.35	67.53	0	307.48	0.69
7	0.1	2520	0	8876.45	0	118.19	0	478.97	0
	0.4	2520	0	8720.34	1.76	118.19	0	468.19	2.25
	0.7	2520	0	8760.17	1.31	118.19	0	503.91	5.21
	1	2520	0	8770.05	1.20	118.19	0	506.95	5.84
10	0.1	3600	0	12621.60	0	168.84	0	707.32	0
	0.4	3600	0	12581.80	0.32	168.84	0	689.52	2.52
	0.7	3600	0	12488.30	1.06	168.84	0	677.07	4.28
	1	3600	0	12452.90	1.34	168.84	0	750.55	6.11

Table L8. Performance With Sharing

Level	Penalty	Allocation		Profit		Demand Loss		Relocation		Transfer	
		Value	Gap (%)	Value ($)	Gap (%)	Value	Gap (%)	Value	Gap (%)	Value	Gap (%)
1	0.1	360	0	1408.69	0	16.88	0	64.58	0	3.29	0
	0.4	360	0	1407.07	0.12	16.88	0	64.63	0.06	3.31	0.61
	0.7	360	0	1402.81	0.42	16.88	0	64.00	0.91	3.31	0.61
	1	360	0	1392.20	1.17	16.88	0	65.21	0.96	3.34	1.52
4	0.1	1440	0	5640.85	0	67.53	0	273.19	0	16.13	0
	0.4	1440	0	5626.29	0.26	67.53	0	275.58	0.88	16.33	1.26
	0.7	1440	0	5605.46	0.63	67.53	0	277.29	1.50	16.04	0.57
	1	1440	0	5591.74	0.87	67.53	0	277.02	1.40	16.14	0.05
7	0.1	2520	0	9893.03	0	118.19	0	423.67	0	24.93	0
	0.4	2520	0	9858.58	0.35	118.19	0	431.29	1.80	26.13	4.80
	0.7	2520	0	9820.18	0.74	118.19	0	428.83	1.22	27.13	8.80
	1	2520	0	9787.94	1.06	118.19	0	432.10	1.99	26.59	6.63
10	0.1	3600	0	14130.90	0	168.84	0	610.08	0	34.91	0
	0.4	3600	0	14071.20	0.42	168.84	0	614.31	0.69	36.10	3.41
	0.7	3600	0	14024.30	0.75	168.84	0	621.88	1.93	38.05	9.00
	1	3600	0	13978.10	1.08	168.84	0	660.63	8.28	37.90	8.57

M Impact of Capacity Sharing When Considering Consumer Demands Change

We note that the capacity-sharing agreement may impact consumer demands in the shared micromobility system. However, to the best of our knowledge, there are no such existing studies. We also note that some studies focusing on the air transport market (e.g., Oum et al. 1996) reveal that airline codesharing can increase the traffic volume in the market. Considering that codesharing is an alternative application of capacity sharing, we follow the findings in Oum et al. (1996) and consider that capacity sharing can increase consumer demands in the shared micromobility system. In this case, capacity sharing can contribute to the shared micromobility system from two perspectives: firstly, by providing firms with opportunities to share their vehicles and, secondly, by increasing consumer demands. Next, we conduct numerical experiments to investigate the overall impact of capacity sharing from both perspectives, as well as its impact solely from the first perspective.

In numerical experiments, we vary the demand level to take into account the impact of capacity sharing on increasing consumer demands. As mentioned in Appendix L, the demand level refers to the extent to which the demand is increased by capacity sharing. With $(\alpha_A, \alpha_B) = (0.1, 0.45)$ and $u = 1.25$, we solve Problem (9) and Problem (13) with the demand level varied from 1 to 2. We set Problem (9), with the demand level fixed at 1, as the benchmark. Then we define

$$\text{No Sharing (\%)} = \frac{\text{the result of Problem (9)} - \text{the result of benchmark}}{\text{the result of benchmark}} \times 100\%,$$

$$\text{Sharing (\%)} = \frac{\text{the result of Problem (13)} - \text{the result of benchmark}}{\text{the result of benchmark}} \times 100\%,$$

$$\text{Increment (\%)(Decrement (\%))} = \text{Sharing (\%)} - \text{No Sharing (\%)},$$

with respect to each instance with a given demand level. Column "Sharing (%)" introduces the overall impact of capacity sharing by increasing demands and providing firms with opportunities to share their vehicles. Column "No Sharing (%)" serves as a reference that shows the impact of increased demands. Column "Increment (%)" (Decrement (%)) offsets the impact of increased demands and shows the impact of capacity sharing solely by providing firms with opportunities to share their vehicles.

Table M9 shows that as the impact of capacity sharing on increasing consumer demands becomes more pronounced, its contribution to profit increments amplifies (see column "Sharing (%)"). Furthermore, the contribution of capacity sharing, solely through offering firms opportu-

Table M9. Impact of Capacity Sharing When Considering Consumer Demands Change

Demand Level	Profit			Relocation			Transfer
	No Sharing (%)	Sharing (%)	Increment (%)	No Sharing (%)	Sharing (%)	Decrement (%)	
1	0	12.67	12.67	0	-13.89	-13.89	3.29
1.25	25.75	41.53	15.79	19.48	3.13	-16.35	4.09
1.5	50.01	69.50	19.49	33.40	15.04	-18.36	5.21
1.75	75.00	95.53	20.53	56.91	36.65	-20.25	6.08
2	99.98	125.81	25.84	68.39	45.25	-23.13	6.99

nities to share their vehicles, also amplifies when demands become larger, as suggested by column "Increment (%)." Similarly, the contribution of capacity sharing in reducing relocation also amplifies when demands grow, as suggested by column "Decrement (%)." We can also observe that the number of vehicles shared between the two firms rises as consumer demands increase.