ABSTRACT

Title of Dissertation: ECONOMETRIC ANALYSIS OF BIKE SHARING SYSTEMS

Huan Cao
Doctor of Philosophy, 2022

Dissertation Directed by: Professor Tunay I. Tunca
Department of Decision, Operations & Information Technologies

I study the efficiency of the dockless bike sharing system, and how to utilize the operational decisions to improve system efficiency and profit.

In the first chapter, I empirically analyze riders' economic incentives in a dockless bike sharing system and explore how to improve the efficiency of this business model. Specifically, I aim to answer three main questions: (i) What is the impact of the number of bicycles in the system on efficiency? (ii) How can bike relocation be best used to improve utilization? (iii) How do the efficiency of dockless and dock-based systems compare? To address these questions, I first build a microeconomic model of user decision-making in a dockless bike sharing system. I then use this model together with transaction-level data from a major dockless bike-sharing firm to structurally estimate the customer utility and demand parameters. Using this estimation in counterfactual analysis, we find that the company can decrease the bicycle fleet size by 40% while maintaining 90% of transactions, leading to estimated savings of $6.5 Million. We further find

that a spatial bicycle rebalancing system based on our customer utility model can improve daily transactions by approximately 19%. Finally, we demonstrate that without bicycle redistribution, a smartly designed dock-based system can significantly outperform a dockless system. Our model provides a utility-based model that allows companies to estimate not only transactions, but also the time and location of lost potential demand, which can be used to make targeted improvements to the geographic bike distribution. It also allows managers to fine-tune bicycle fleet sizes and spatial rebalancing parameters. Further, our structural demand modeling can be used to improve the efficiency of dock-based systems by helping with targeted dock location decisions.

In the second chapter, using data from a major dockless bike sharing system in Beijing, I study the subscription behavior and its relationship with the service level and price. Specifically, I develop an econometric model to study the subscription behavior for both existing subscribers and new sign-ups, respectively, and build up a functional relationship between the service level and system demand level and reveal the dynamics and interplays among subscription, system demand, and service level, which helps to recover the evolution of the number of subscribers over time. Based on all the estimation results and functional relationships, I then construct an empirical framework and straighten out the relationship between company profit and bicycle fleet size/subscription price. The counterfactual results show that the marginal benefit of deploying a larger bicycle fleet is decreasing, and the company should be cautious in determining and adjusting the bicycle fleet size. In addition, the examination demonstrates that the current price is too low, and raising the price properly can achieve about a 25% profit increase. The results also show the value of understanding riders' sensitivity to price and how to use it to better their operational decisions and accomplish better financial results. The counterfactual framework I proposed can be utilized in various policy evaluations and provides important insights and recommendations to the bikeshare companies and regulators.

ECONOMETRIC ANALYSIS OF BIKE SHARING SYSTEMS

by

Huan Cao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Tunay Tunca, Chair/Advisor
Professor Wedad Elmaghraby
Professor Michael Fu
Professor Ginger Zhe Jin
Professor Yi Xu

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Tunay Tunca for giving me an invaluable opportunity to work on challenging and extremely interesting projects over the past five years. He has always made himself available for help and advice and there has never been an occasion when I've knocked on his door and he hasn't given me time. It has been a pleasure to work with and learn from such an extraordinary individual.

I would also like to thank my coauthor, Professor Weiming Zhu. Without his encouragement and help, this thesis would have been a distant dream. Thanks are due to Professor Wedad Elmaghraby, Professor Michael Fu, Professor Ginger Zhe Jin, and Professor Yi Xu for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript.

I would like to thank Jianfeng Lin, Xingliang Zheng, and Meituan's AI team for providing us the comprehensive datasets as well as collaboration opportunities. Also, I thank Rushi Ye for helping me get access to the company and obtain data from them.

My colleagues at Mobike/Meituan Bike have enriched my industry collaboration experience in many ways and deserve a special mention. Jiao Du and Anqi Fu helped me start off by introducing the technology tools and dashboards used at Mobike. Yu Wang and Nan Tang provided clarifications for some complex metrics and rules. My interaction with Xiaoxu Zhao and

Xinlei Hu has been very fruitful throughout our experimentation project.

I owe my deepest thanks to my family - my mother and father who have always stood by me and guided me through my career, and have pulled me through against impossible odds at times. Words cannot express the gratitude I owe them. Also, my special thanks go to my girlfriend, Shijia Liu. Her understanding and support are what encourage me to finish the dissertation smoothly.

It is impossible to remember all, and I apologize to those I've inadvertently left out.

Table of Contents

Table of Contents ... iv

List of Tables ... vi

List of Figures ... vii

Chapter 1: Introduction ... 1
- 1.1 An Overview of Bike Sharing Systems ... 1

List of Abbreviations ... 1

Chapter 2: Extracting Efficiency from Chaos: Rider Behavior, Design and Performance of Dockless Bike Sharing Systems ... 3
- 2.1 Introduction ... 3
- 2.2 Literature Review ... 7
- 2.3 Data Description ... 9
- 2.4 Model ... 13
  - 2.4.1 Rider Arrival Rates ... 13
  - 2.4.2 Rider Conversion Rates ... 14
- 2.5 Structural Estimation ... 19
  - 2.5.1 Rider Origins ... 19
  - 2.5.2 Bike-to-Rider Distance Distribution ... 20
  - 2.5.3 Estimation Results ... 23
- 2.6 Counterfactual Analysis ... 25
  - 2.6.1 Setup ... 26
  - 2.6.2 Bike Fleet Size ... 30
  - 2.6.3 Bike Redistribution Strategy ... 32
  - 2.6.4 Dockless vs. Docked Bikesharing ... 35
- 2.7 Conclusions ... 47

Chapter 3: Improving Profit through Improving System Performance in Bike Sharing ... 50
- 3.1 Introduction ... 50
- 3.2 Literature Review ... 52
- 3.3 Empirical Setting and Data ... 54
  - 3.3.1 Empirical Setting: The Meituan Bike System ... 54
  - 3.3.2 Meituan Bike Data: Subscription Behavior and Service Level ... 55
  - 3.3.3 Subscription Behavior ... 56

3.3.4 Service Level 57
3.4 Econometric Specifications and Results 58
3.4.1 Subscription Renewals 58
3.4.2 New Sign-ups 62
3.5 Counterfactual Analysis 65
3.5.1 Setup 66
3.5.2 Bicycle Fleet Size 68
3.5.3 Price 70
3.6 Concluding Remarks 71

Bibliography 74

List of Tables

2.1 Summary Statistics 10
2.2 Example of App-opening Data 12
2.3 Example of POI Data 13
2.4 Estimation Results 25
2.5 Example of Trip Information 28
2.6 Performance Comparison between Different Redistribution Strategies 34
2.7 Estimation Results 41

3.1 Summary Statistics 57
3.2 Estimation Results of Renewal Events 61
3.3 Estimation Results of Auto-renewal 63
3.4 Estimation Results of New Sign-up Events 64

List of Figures

2.1 Region of Interest in the City of Changsha 9
2.2 Bike Trip Summary. Panels (a), (b) and (c) present the distributions for trip duration, trip distance and trip starting times for all bike trips for Mobike users in Changsha initiated on June 13th, 2019 between 7 AM and 10 AM. 11
2.3 Panel (a) Logarithm of the Number of Bikes in Each Block at 7:00 AM. Panel (b) Difference of the Logarithm of the Number of Bikes in Each Block between 7:00 AM and 9:00 AM 12
2.4 Example of a Square: Each Blue Balloon in the Square Represents a POI 21
2.5 The impact of the number of bikes on rider conversion rates and the bike utilization. 31
2.6 Transaction Performance Ratio between Dockless and Dock-based Systems . . . 42
2.7 The Number of Bicycles in the Top 5% Low-Demand Blocks 44
2.8 Bike Imbalance Index Comparisons 46

3.1 Number of Existing Subscribers and New Sign-ups 56
3.2 Service Level vs BA_ratio 67
3.3 The impact of fleet size on company revenue and service level. 70
3.4 The impact of fleet size on company revenue and service level. 71

Chapter 1: Introduction

1.1 An Overview of Bike Sharing Systems

While enjoying the fruits of economic growth and development, people are also suffering from the side-effects. For instance, traffic congestion and poor air quality have been serious problems in urban cities across the world for many years, and passenger vehicles are the major culprit: 80% of air pollution in some Asian cities can be directly attributed to private passenger vehicles (World bank and Asian Development Bank Report, 2012). Traffic congestion itself is also detrimental. According to the Texas Transportation Institute (TTI), an average commuter in U.S. urban areas loses 34 hours and $750 each year because of traffic delays. From this perspective, bike-sharing systems can help mitigate the pressure and provide a convenient and affordable means of transport for urban residents.

In recent years, urban bike-sharing systems developed in two main forms: *Dock-based* systems that rely on fixed location docks to park at and pick up bicycles from, and *Dockless* systems that allow users to pick up and drop off bikes anywhere. By eliminating reliance on fixed docks, dockless systems provide significantly improved mobility and convenience to riders, however, they have downsides such as urban clutter and bike pile-ups. As a result, there is an ongoing debate about which business model works better and how to improve the system efficiency and sustainability. I study both short-term and long-term questions related to dockless bike sharing systems. In Chapter 2, I study the dynamics and evolution of the dockless bike sharing system, analyze ways to improve system efficiency, and compare the two different bike sharing models. In Chapter 3, I study how service level and subscription fee affect subscription behavior and how to utilize these relationships to optimize operational decisions and maximize company profit.

Chapter 2: Extracting Efficiency from Chaos: Rider Behavior, Design and Performance of Dockless Bike Sharing Systems

2.1 Introduction

Bike sharing as a convenient and affordable means of transport has gained increasing popularity in recent years. The shared use of bicycle fleet mitigates the vehicle emissions and reduces congestion in the city. Moreover, biking can be an effective 'first and last mile' (i.e., the first and final leg of a commuter's journey) urban mobility solution that enhances connectivity to public transit. Bike sharing system has been through several waves of transformation (DeMaio [1]) and has been growing rapidly across the world in the past decade with the advent of smart phone. In 2019, shared bikes, e-bikes and scooters amount to 136 million trips in the United States, reporting 60% growth from 2018 (NACTO 2020 [2]).

A typical dock-based bike sharing system includes a communal stock of bikes distributed over a network of parking stations. Each station provides a number of parking docks. Bike users can unlock available bikes from stations and return them to stations with vacant docks. However, potential riders can be discouraged from using such dock-based service due to the lack of available bikes or docks in the proximity. Even the most established bike sharing systems may not be able to offer a satisfactory service level during peak hours: users in seven different

New York City neighborhoods have a 25% or higher chance of finding a full station at peak hour (PYMNTS 2017 [3]), mostly due to the fixed and inflexible dock locations and capacities of such stations.

To inject flexibility into the system, the dockless sharing scheme has emerged from China. Mobike (later acquired by Meituan and renamed to Meituan Bike) is one of the largest dockless bike sharing companies in China. Present in more than 180 cities, Mobike owns more than 7 million bikes and serves 200 million registered users on a daily basis. The major modification of the dockless system is the "park anywhere" policy, which gets rid of the concept of stations and allows bikes to be parked anywhere in the city so long as it does not obstruct the traffic. Different from the some well-studied bike share companies (e.g., Citi Bike, Lime) whose bikes are parked at pre-installed stations, Mobike is completely dockless and can be parked at most of the places in a city. As a result, most of the bike-searching, and all the bike-unlocking processes reply on the use of the mobile application.

By the end of 2019, a number of companies have adopted the dockless bike sharing system and have served more than 360 Chinese cities with 47 million daily trips (The State Information Center 2019 [4]). While dockless system has gained tremendous popularity in China, the opposite has been true in the U.S. The influx of dockless sharing bikes across the U.S. did not translate into substantial mobility gains. In year 2017, dockless bikes made up 44 percent of all shared bikes on the street, yet they accounted for a mere 4 percent of the bike-share trips (NACTO 2018 [5]).

In this paper, utilizing the comprehensive datasets obtained from Mobike, we aim to shed light on two streams of questioning. The first is related to the efficiency of the dockless bike sharing system. To maintain a high rider conversion rate and sustain a satisfactory performance level, dockless bike sharing companies normally oversupply bikes in their operating regions. We are curious about how efficient is the dockless system with different bike fleet size and how can the company maintain the system service level by some operational interventions. Second, although the dockless system seems to be more appealing to riders for its “park anywhere” policy, the bike locations are more manageable in the dock-based system. Therefore, from system design’s perspective, which bike sharing scheme should a bike sharing firm go for? Comparing to the dock-based system, does the dockless system always perform better?

To address these questions, we empirically explore rider decision making process and efficiency of the dockless bike sharing system. We develop a novel framework to model the rider decision process by explicitly accounting the rider arrival and the rider-to-bike distance. Using transactional data from Mobike, we estimate the usage drivers for the system. Based on the estimation results, we then run counterfactual analysis to demonstrate how the number of deployed bikes affect the performance of the system. We also study measures to improve efficiency, and compare the effectiveness of dockless versus dock-based systems.

The main contributions of the paper is threefold.

1. Methodology. The dockless bike sharing system differs from the dock-based one in that the bike can be anywhere in the operating region. For this reason, the state space of each bike is infinite and it is infeasible to adopt the classic demand estimation framework developed for the dock-based system. In this paper, we propose a novel approach that construct the distribution of the bike-to-rider distance to model the rider decision process. Our model provides a utility based model that allows companies to estimate not only transactions, but also time and location of lost potential demand, which can be used to make targeted

improvements to the geographic bike distribution. The method can be extended to studying demand for other dockless micromobility systems.

2. Practical guidance. Our study provides important practical guidance for Mobike to evaluate the bicycle fleet size and the bike redistribution strategy. Our utility based model allows managers to fine tune fleet sizes and spatial rebalancing parameters. The framework can also be easily applied to other dockless bike sharing systems in other cities to recover potential demand and evaluate operational policies.

3. General insights. Our analysis highlights the advantages and disadvantages of the dockless system and the dock-based system. We provide strong evidence to show that the dockless system tends to attract more riders because of the flexible parking policy. At the same time, the dock-based system has an edge on disciplining rider behaviors by having a network of stations, which imposes better control on the bike distribution of the system. Therefore, everything the same, the dockless system is able to achieve better performance than its docked counterpart, yet it requires constant bike redistributions to remain effective in the long run.

The remainder of this paper is organized as follows. Section 2.2 provides the literature review. Section 2.3 describes the data we have. Section 2.4 introduces the model. Section 2.5 presents our empirical estimation of this model. Section 2.6 discusses the counterfactual analysis. Section 2.7 provides our concluding remarks.

2.2 Literature Review

Our paper estimates the rider demand in a dockless bike sharing system and thus contributes to the literature on demand estimation in bike sharing system. Previous papers have come up with several approaches to estimate rider demand. Singhvi *et al.* [6] and El-Assi *et al.* [7] obtain demand prediction through observed bike transaction data, and investigate how factors such as weather condition, zonal level socio-demographic characteristics affect bike demand. Goh *et al.* [8] propose a rank-based choice model to reveal the primary rider demand of the bike sharing system. Mellou and Jaillet [9] develop a linear programming framework to identify the shifted demand resulting from station outages. Our paper adopts a random utility model and constructs rider's utility gain from using the bike sharing service. Combining with our datasets, we recover key customer preference parameters and the uncensored demand generating process. Closest to our work are Kabra *et al.* [10] and He *et al.* [11], which model the rider's choice and structurally probe how the density of bike station affects bike usage. Specifically, Kabra *et al.* [10] estimate the impact of the walking distance between customers and bike stations on bike usage, and emphasize the need for dense station networks. He *et al.* [11] estimate origin-destination paired demand, and conduct counterfactual analysis to provide insights on network design and station expansion strategy.

All the above-mentioned works consider dock-based bike sharing systems, where riders must access a docking station to unlock or return their bikes. In contrast, our paper is among the first to study the dockless bike-sharing system. The fact that there is no station in the dockless system imposes an unique estimation challenge that makes the previous estimation approaches not applicable. To be specific, unlike previous literature on docked bike sharing system that applies the BLP model (Berry *et al.* [12]) by defining each bike station a distinct product (Kabra *et al.* [10]), or each origin and destination station pair as a product (He *et al.* [11]), it is computationally not tractable to treat individual bike as a product due to the enormous amount of bikes (87,040 bikes in total) and their time-varying features (e.g., location and availability of each bike). In our paper, we develop a novel estimation framework to recover the uncensored rider demand in the dockless bike sharing system. Specifically, we derive the distribution of the bike-to-rider distance to facilitate the estimation of the rider conversion rate. Meanwhile, we leverage two unique datasets that document the Mobike application opening data and the Point of Interest (POI) to construct customer arrival rate. Linking the conversion rate and the constructed arrival rate to the actual transaction data, we then recover the uncensored demand.

Several recent papers study the welfare generation and the economic implications of the dockless bike sharing system. Cao *et al.* [13] examine the dockless bike sharing market and demonstrate that when a platform enters a new market that is dominated by another platform, the entry expands the market for the incumbent, increasing its total demand as well as the revenue per trip. Chu *et al.* [14] show that the entry of bike sharing leads to a 29% per kilometer reduction in the price premium for housing close to metro station. In a similar setup, Cui *et al.* [15] demonstrate that, responding to the entry of bike sharing system, restaurants reduce prices by 1.4% and boost their service quality by 0.7%.

By examining the impact of bike distribution on bike demand through counterfactual analysis, our work also connects to the theoretical works in bike-sharing that focus on the bike rebalancing problem. In a dock-based bike sharing system, each rider takes a bike from a station and parks it at another station that is close to her destination. Some stations then become full while others empty. Hence, bike rebalancing is usually required to redistribute bikes among stations

(a) Region of Interest in the City of Changsha (b) Classification of Blocks within the Region of Interest

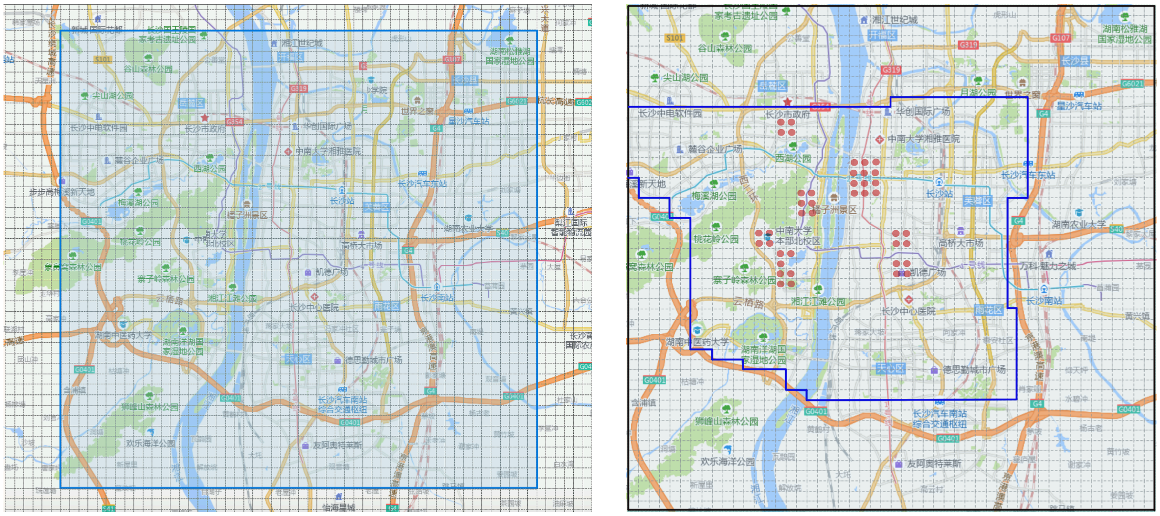

Figure 2.1: Region of Interest in the City of Changsha

to better match supply with future demand. Specifically, Chemla *et al.* [16] and Dell'Amico *et al.* [17] examine a static rebalancing problem, in which bikes need to be repositioned given a fixed bike distribution to better match supply with demand. More recently, O'Mahony *et al.* [18], Schuijbroek *et al.* [19] and Ghosh *et al.* [20] investigate a dynamic rebalancing problem where bikes need to be rebalanced continuously, especially during periods of high demand.

2.3 Data Description

For our empirical analysis, we obtain data from Mobike on trip information, bike location as well as App-opening records. Our datasets cover trips starting within the main operating region in the city of Changsha, a large city in central China with its population topping 7 million. Panel (a) of Figure 2.1 highlights the region of interest: an area encompassed in the square that is about 25 kilometers by 26 kilometers at the center of the city. We classify the main operating area in

Table 2.1: Summary Statistics

Each Transaction	Mean	St. Dev.	Min	Max
Duration (min)	11.29	14.96	1.00	395.00
Distance (km)	1.25	1.20	0.00	25.50

Changsha into three categories, namely uptown, midtown and downtown. Specifically, the red dots shown in Panel (b) of Figure 2.1 represent the downtown areas, which are typically where the main shopping areas and the college locate at, the rest of the area encompassed by the blue boundary line is midtown, and finally, the outer belt between the blue and the black borderline is classified as uptown. The geolocation information will later be incorporated in our empirical analysis. In what follows, we introduce the main datasets we use for our study.

Trip Information The first dataset contains detailed information about all trips from Mobike in Changsha on June 13th, 2019 during the morning rush hour (7:00 AM to 10:00 AM). For each trip, we observe the starting and ending time as well as the longitude and latitude for the corresponding pick-up and drop-off locations. There are 11,642 trips in total. Table 3.1 displays the summary statistics, and Figure 2.2 presents the empirical distributions of trip duration, distance and starting time. Note that around 95% of the trips are within 30 minutes, and over 98% of the trips are within 5 kilometers.

Bike Location Snapshots The snapshots data documents the by-hour location of each bike which is sent from the GPS embedded in the bike. As not all available bikes in Changsha register a transaction on a daily basis, and those with transaction data can be relocated by the company to better balance the demand and supply, the bike snapshots dataset augments our transaction data by providing additional information on the bike availability in different operating regions in each time window.

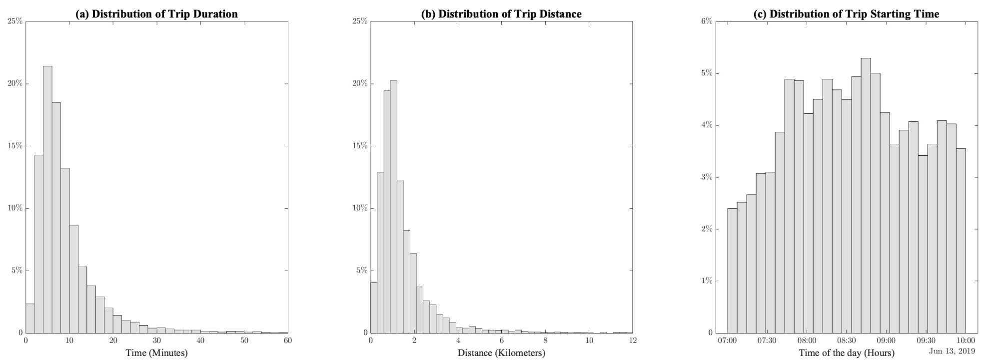

Figure 2.2: Bike Trip Summary. Panels (a), (b) and (c) present the distributions for trip duration, trip distance and trip starting times for all bike trips for Mobike users in Changsha initiated on June 13th, 2019 between 7 AM and 10 AM.

We divide the operating area of the bike share system into 2,500 blocks (approximately 500 by 500 meter). Then, utilizing the hourly snapshot data, we obtain the number of available bikes in each block for each 10 minute time interval. To do so, we count the number of bikes in each block from the bike snapshots data, and initialize the number of bikes at the beginning of each hour. Next, by combining the initial locations with the trip data, which traces the inflow and outflow of bikes, we update the number of bikes in each block and for each 10 minute interval as

$$n_{j,t+1} = n_{jt} + n_{jt}^{\text{inflow}} - n_{jt}^{\text{outflow}}, \tag{2.1}$$

where n_{jt} is the number of bikes in block j at time interval t, n_{jt}^{inflow} and n_{jt}^{outflow} are the total number of bikes coming in and leaving block j during time window t, respectively. Panel (a) of Figure 2.3 presents the available bikes, measured as the logarithm of the idle bikes, in each block at 7:00 AM, during which bikes are mostly concentrated at downtown areas and university campuses - places of high demand during the morning peak hours. In the meantime, we present the difference of the logarithm of the number of available bikes between 7:00 AM and 9:00 AM

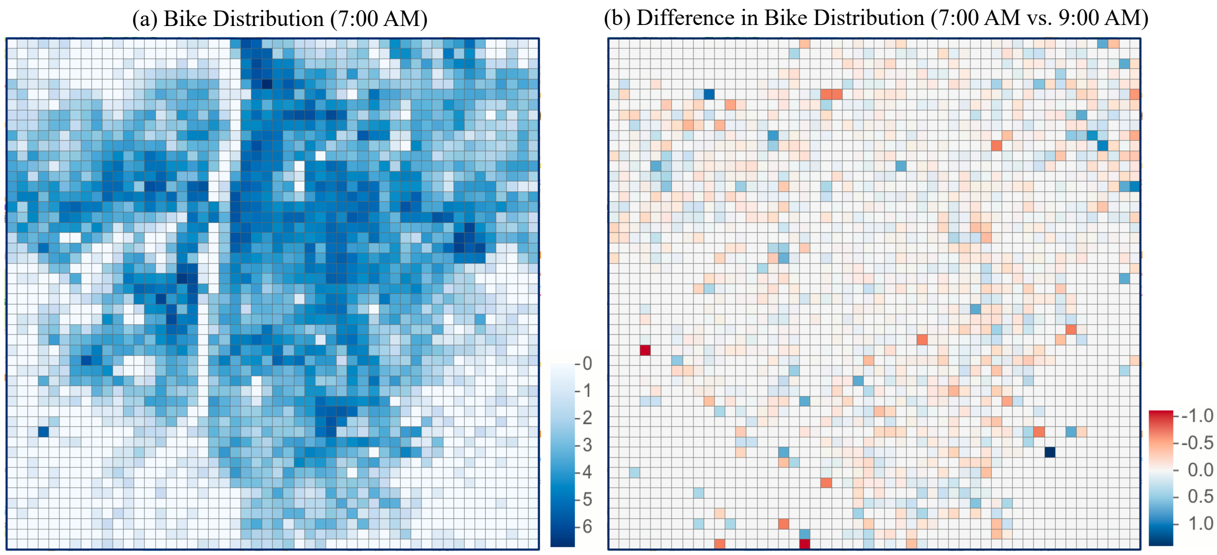

Figure 2.3: Panel (a) Logarithm of the Number of Bikes in Each Block at 7:00 AM. Panel (b) Difference of the Logarithm of the Number of Bikes in Each Block between 7:00 AM and 9:00 AM

in Panel (b), in which blocks in darker blue are popular destinations, whereas blocks in dark red are popular trip origins.

App-opening Data We also observe the app opening behavior of the Mobike users. Every time a user opens the Mobike system, either through Mobike Application or Wechat's HTML5 mini-program, the user's operations and location (up to the precision of GPS in the phone) will be stored. The example shown in Table 2.2 presents five of such app-opening instances.

Table 2.2: Example of App-opening Data

User ID	DateTime	Longitude	Latitude	Source
10001	2019-05-29 18:33:47	113.03135	28.1765	Mobike Application
10001	2019-05-29 18:34:14	113.03131	28.17623	Mobike Application
10001	2019-05-29 18:37:04	113.03172	28.17627	Mobike Application
10002	2019-05-29 07:37:55	113.02499	28.14659	Wechat
10003	2019-05-29 11:58:18	112.98465	28.26349	Mobike Application

Points of Interest: In addition to bike- and app-related data, we collect points of interest (POI) data in Changsha using AutoNavi (Gaode Map) API. The POI dataset includes locations where Mobike users tend to accumulate and is classified into five different categories, namely restaurants, stores, transportation, residence, educational institution. In particular, this dataset contains the geographic coordinates and category for around 200,000 points in the operating area. Sample data is presented in Table 2.3.

Table 2.3: Example of POI Data

Name	Longitude	Latitude	Type
Wenchangge	112.9784	28.21317	Transportation
Luxiyuan	112.9391	28.19775	Residence
Xinxin Pizza	112.9260	28.15971	Food
School of architecture, Hunan University	112.9479	28.17895	School

2.4 Model

To study the effectiveness of a dockless bike share system, we build a structural model to capture the rider's decision making process. The model consists of two components: (i) the *uncensored* arrival rate of riders in each block, and (ii) a potential rider's decision to choose between Mobike and other commuting options given the location of available bikes in her vicinity. Using these two components, we can obtain the expected number of completed transactions at acertain time and location. We next explain these two components in detail.

2.4.1 Rider Arrival Rates

We start by analyzing the customer journey for Mobike users to facilitate the derivation of the regional demand over each 10-minute time window. When a Mobike user is to commute, she (subsequently, we will refer to the rider by the female pronoun) could spot the bike either through eyeballing the neighborhood, in which case there is no need to access the application, or through checking the Mobile application to identify available bikes nearby. If the rider in the former case opts to commute through means other than Mobike, her searching behavior will not be recorded by the Mobike application. At the same time, not all the app-opening records are generated from users' intention to search for a bike. Mobike users may also open the app to check their subscription status or past trips. Consequently, using the app-opening instances directly as the number of arrived users can lead to biased demand estimates. That said, the app-opening data still serves as a good proxy for the actual demand, as converted users in the former case and all the users in the latter case are captured by the app-opening data. Hence, we parametrize the potential riders that originate from block j in time window t, denoted as λ_{jt}, by the following expression:

$$\lambda_{jt} = \alpha_0 + \alpha_1 \cdot \lambda_{jt}^{app} + \boldsymbol{\beta} \cdot \mathbf{X}_{jt}, \tag{2.2}$$

where λ_{jt}^{app} represents the number of app-opening events observed during time window t in block j, $\mathbf{X}_{jt}$ contains hour-of-the-day dummies and $\mathbf{N}_j^{poi}$, i.e., the number of POI of each category in block j. The uncensored demand rate λ_{jt} is shaped by α_0, α_1 and $\boldsymbol{\beta}$, whose value will be jointly estimated from the full estimation model later in Section 2.5.3.

2.4.2 Rider Conversion Rates

The second part of our model constructs the rider utility from commuting through Mobike, based on which we model the consumer's decision-making process. For a potential commuter i who spawns in block j during the time window t, the utility she would gain from using Mobike is

$$U_{ijt} = a_{ij} - dis_{ijt} \, , \tag{2.3}$$

where a_{ij} captures the intrinsic value of a trip initiated from block j rider i obtains using Mobike, dis_{ijt} denotes the walking distance from rider i to the her nearest bike in block j at time window t, and the coefficient of dis_{ijt} is normalized to 1 without loss of generality. This way, a_{ij} can be interpreted as the maximum distance the rider is willing to walk to pick up a bike from Mobike. In reality, bikes are not evenly distributed across the city as illustrated by Figure 2.3, and downtown areas usually have more alternative means of transportations readily available for the rider to choose from. Thus, the same rider may set location-dependent expectation on the walking distance. Additionally, rider's willing-to-walk can also vary with the upcoming trip distance. Users may be willing to walk farther to pick up a bike if they need to ride for a longer distance. Therefore, we write a_{ij} as

$$a_{ij} = a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i \, , \tag{2.4}$$

where $a_{g(j)}$ captures the potential geographic region fixed effect and $g(j) : j \to \{U, M, D\}$ maps each block j to one of the three geographical regions: uptown, midtown and downtown as marked in Panel (b) of Figure 2.1. Meanwhile, a_2 incorporates the effect of trip distance on the rider's valuation from the trip. To elaborate, we divide the trips into two groups, with short riding distance group denoted as "S" and long riding distance group "L", and we use $\mathbf{1}_{\{i \in L\}}$ to indicate if the trip from rider i is a long distance one. This way, the coefficient a_2 is the additional value the rider perceives from the long distance trip. Finally, the error term ϵ_i is assumed to follow a distribution such that the willingness-to-walk is Beta distributed with mean of $a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}}$ and unknown variance σ^2, which captures all remaining heterogeneity in the value brought by the trip. Combining Equation (2.3) and (2.4), rider's utility gain from using Mobike for a specific trip can be written as:

$$U_{ijt} = a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} - dis_{ijt} + \epsilon_i \,, \tag{2.5}$$

where the set of parameters to be estimated are $(a_U,\ a_M,\ a_D,\ a_2,\ \sigma^2)$. The utility in Equation (2.5) is measured relative to the outside option, which is normalized to 0. Notably, the outside option encompasses all the alternative means of transportation, including other bike-share services offered by Mobike's competitors. The rider chooses to use Mobike if $U_{ijt} \geq 0$. Thus, rider i's conversion probability would be:

$$\mathbb{P}_{ijt} \equiv Pr\{U_{ijt} \geq 0\} = Pr\{dis_{ijt} \leq a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i\} \,. \tag{2.6}$$

In Equation (2.6), dis_{ijt} represents the walking distance for rider i to the nearest bike that locates in block j; $a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i$ is potential rider i's willingness-to-walk in block j at time period t.

Note that we assume $a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i$ follows log-normal distribution. Therefore, all the commuters, even for those who have near zero willingness-to-walk for sharing bikes (since they prefer other commuting tools), are captured by the model. Specifically, our demand term λ_{jt} captures the arrival time and location of all the potential riders in the market, and the conversion probability $\mathbb{P}_{ijt}$ describes how likely each commuter will convert and choose to use the sharing bike service.

Unlike dock-based systems in which bikes are parked at one of the bike stations, bikes in dockless systems can be parked almost anywhere on the street which makes the state space for the bike location considerably larger than the dock-based case. Therefore, instead of calculating the distance between available stations and possible rider locations (He *et al.* [11], Kabra *et al.* [10]), we propose a novel approach to keep track of the location for not only the potential rider, but all the bikes in the neighborhood of the rider so as to identify the distance the rider need to walk to get to the nearest bike. Notably, we do observe in our dataset hourly snapshots of bike locations, yet the data cannot be directly applied to calculate the shortest bike-to-rider distance for three reasons. First, the accuracy of bike GPS is not perfect. According to Mobike, around 80 percent of the GPS data provide accuracy within 100 meters. Second, unlike the dock-based system for which available bikes are locked on fixed docks, it is possible in the dockless system to move a bike around without unlocking it. However, this type of movement may not be captured until Mobike updates the GPS information from the bike. Third, to keep track of every single bike and update the distance whenever a transaction initiated or finished is computationally burdensome. Therefore, instead of calculating the exact bike-to-rider distance, we construct the distribution of the bike-to-rider distance using the number of bikes and the bike location snapshots in each block which we describe next.

To derive the bike-to-rider distance distribution, we assume that riders only consider bikes located within their blocks. In doing so, we are able to treat each block as a separate market. For a given block j, we denote the number of bikes at the beginning of time window t as n_{jt}. Meanwhile, we have formulated the uncensored rider arrival rate in block j during time window t as λ_{jt} through the number of app-opening events. However, the block level arrival rate does not pinpoint the exact location where each rider originates. To provide more accurate estimates on customer origins, we divide each block j into m_j potential squares (the details of the division are presented in Section 2.5.1). We assume that riders who arrive at block j can spawn at one of the squares o with probability p_{oj}, where $\sum_{o=1}^{m_j} p_{oj} = 1$. Specifically, the value of each p_{oj} ties to the number of POI around location o and is derived as an outcome from our structural estimation, which we discuss in greater details in Section 2.5.1. Additionally, we assume that all the bikes are independently and identically distributed (i.i.d.), which implies that the Manhattan distances between the bikes and origin o are also i.i.d. We denote the distance between origin o and bike k in block j as x_{ojk}. The p.d.f. and c.d.f. for x_{ojk} are $f_{x_{oj}}(\cdot)$ and $F_{x_{oj}}(\cdot)$, respectively. Therefore, the distance between point o and its nearest bike can be expressed as $z_{oj} = \min_{k=1,\ldots,n_{jt}}\{x_{ojk}\}$, which has the following c.d.f.:

$$
\begin{aligned}
F_{z_{oj}}(z|n_{jt}) &= Pr(z_o \leq z) \\
&= Pr(\min_{k=1,\ldots,n_{jt}}\{x_{ojk}\} \leq z) \\
&= 1 - \left(1 - F_{x_{oj}}(z)\right)^{n_{jt}}. \qquad (2.7)
\end{aligned}
$$

Subsequently, when block j contains n_{jt} bikes at time t, the c.d.f. of the shortest bike-to-rider distance can then be written as the weighted sum of the c.d.f. of the shortest bike-to-rider distance from each potential rider origin point:

$$
F_{dis_{jt}}(x|n_{jt}) = \sum_{o=1}^{m_j} p_{oj} \times F_{z_{oj}}(x|n_{jt}). \qquad (2.8)
$$

Meanwhile, we can rewrite rider i's conversion rate specified in Equation (2.6) as:

$$
\mathbb{P}_{ijt} = F_{dis_{jt}}(a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i | n_{jt}), \qquad (2.9)
$$

which we use to derive the average conversion rate of block j in time window t in the next section.

2.5 Structural Estimation

In this section, we derive the rider origins and, based on which, derive the bike-to-rider distribution. We then apply nonlinear least squares method to recover the parameters in Equations (2.2) and (2.5). In our analysis, we divide the morning rush hour (7 AM to 10 AM) into 10-minute time windows, and consider each window as our unit of observation. This leaves us with $T = 18$ time windows in our data.

2.5.1 Rider Origins

To derive the distribution of the rider origins, we divide each block into 50-meter-by-50-meter squares (each block then contains around 100 squares), and, if a potential rider arrives at a specific square o, we assume that rider is equally likely to appear at each point in that square. Notably, we do observe the longitude and the latitude when a customer opens the Mobike application, yet the app-opening data cannot be readily used as rider's origination. This is because most users only open their Mobike application to unlock the bike next to them, using the app-opening data as the proxy for customers' origins can severely bias the estimation of customers' willingness to walk.

Next, we derive p_{oj}, the probability that a customer who shows up in block j belongs to square o, by leveraging the POI information in block j. By definition, POI are defined as places where customers tend show up around. Yang *et al.* [21] indicate that the number of POI in a region can be highly correlated to the number of trips generated, and Mobike uses POI as one of the key predictors for forecasting customer demand. In our case, certain types of POI such as residential areas could generate more demand during the morning rush hours than the other POI (e.g., restaurant), we thus assign demand weight to each POI type as: $\mathbf{W} = (w_{Resid}, w_{Retail}, w_{Rest}, w_{Other})$. And for each potential rider origin o in block j, its total demand weight w_{oj} is calculated as

$$w_{oj}(\mathbf{W}) = \mathbf{W} \cdot \mathbf{N}_{oj}^{poi}\,, \tag{2.10}$$

where $\mathbf{N}_{oj}^{poi}$ is a column vector containing the number of POI of each category in square o in block j. We also add a constant 1 as a linear transformation to make sure the weight is non-zero for squares that contain zero POI. Figure 2.4 shows a sample square with all the POI (the five blue balloons) inside. The number next to each blue balloon is the assigned weight, and the demand weight for the center point of the square, o, is the sum of all the POI's weights and is equal to 22 in the example.

With the aggregated weights for all the possible origins in block j, we can estimate p_{oj} by normalizing the weights as

$$p_{oj}(\mathbf{W}) = \frac{e^{w_{oj}(\mathbf{W})}}{\sum_{u=1}^{m_j} e^{w_{uj}(\mathbf{W})}}\,, \tag{2.11}$$

where m_j is the number of squares in block j. This way, the arrival rate for square o of block j at time window t can be expressed as $\lambda_{jt} \cdot p_{oj}(\mathbf{W})$.

2.5.2 Bike-to-Rider Distance Distribution

With the rider origination derived, we now develop closed-form expressions for the distribution of the bike-to-rider distance and the average conversion rate in block j. As there are in total

Figure 2.4: Example of a Square: Each Blue Balloon in the Square Represents a POI

53,960 bikes from Mobike in Changsha during the timespan of our data, we assume that there is at least one bike in each block. Given that the distance each rider needs to walk is bounded by the size of the block, we assume that x_{oj} follows Beta distribution:

$$f_{x_{oj}}(x) = \frac{x^{b_{1oj}-1}(1-x)^{b_{2oj}-1}}{B(b_{1oj}, b_{2oj})} \equiv f(x|b_{1oj}, b_{2oj}), \tag{2.12}$$

where $x \in (0,1)$, $B(b_{1oj}, b_{2oj}) = \frac{\Gamma(b_{1oj})\Gamma(b_{2oj})}{\Gamma(b_{1oj}+b_{2oj})}$ with $\Gamma(\cdot)$ being the Gamma function, and b_{1oj} and b_{2oj} dictate the shape of the distribution for each square o in each block j and are to be estimated. Specifically, to estimate shape parameters for each potential origin in each block, we use the hourly bike snapshot data during morning rush hour (7:00 AM-10:00 AM), from June 7th to 20th, 2019. We treat the bike coordinates in each block as sample bike locations whose distance to each rider origin is generated from an origin-specific beta distribution. For all the bikes located in block j as according to the snapshot data, we obtain the origin-bike distance by using the Manhattan distance metric and distance formula in Veness [22]: Distance $= R \cdot (c_{Lat} + c_{Lng})$, where $R = 6371$ Kilometers which is the radii of the earth. In addition, c_{Lat} and c_{Lng} are calculated as:

$$c_{Lat} = 2 \cdot \arctan\left(\sqrt{\frac{\sin^2(|\text{DLA} - \text{SLA}| \cdot \pi/360)}{1 - \sin^2(|\text{DLA} - \text{SLA}| \cdot \pi/360)}}\right), \tag{2.13}$$

$$c_{Lng} = 2 \cdot \arctan\left(\sqrt{\frac{\sin^2(|\text{DLO} - \text{SLO}| \cdot \pi/360) \cdot \cos(\text{DLA} \cdot \pi/180) \cdot \cos(\text{SLA} \cdot \pi/180)}{1 - \sin^2(|\text{DLO} - \text{SLO}|) \cdot \pi/360 \cdot \cos(\text{DLA} \cdot \pi/180) \cdot \cos(\text{SLA} \cdot \pi/180)}}\right),$$

where SLA and DLA are abbreviations for starting and destination latitudes, and similarly, SLO and DLO represent the starting and destination longitudes for the trip. We then use the distances to fit the Beta distribution, $F_{x_{oj}}(\cdot)$, and estimate b_{1oj} and b_{2oj} for each origin o. Plugging in b_{1oj} and b_{2oj} to Equation (2.7), we have:

$$\begin{aligned} F_{z_{oj}}(z|n_{jt}) =& 1 - \left(1 - F_{x_{oj}}(z)\right)^{n_{jt}} \\ =& 1 - (1 - F(z|b_{1oj}, b_{2oj}))^{n_{jt}} \equiv F_z(z|n_{jt}, b_{1oj}, b_{2oj}). \end{aligned} \tag{2.14}$$

In doing so, we can derive the conversion probability of rider i in each block j during time window t, $\mathbb{P}_{ijt}$, by plugging the bike-to-rider distribution for potential origin o, $F(\cdot|b_{1oj}, b_{2oj})$, and its corresponding probability p_{oj} into Equation (2.8):

$$F_{dis_{jt}}(x|n_{jt}, b_{1oj}, b_{2oj}, \mathbf{W}) = \sum_{o=1}^{m_j} p_{oj}(\mathbf{W}) \times F_{z_{oj}}(x|n_{jt}, b_{1oj}, b_{2oj}). \tag{2.15}$$

Finally, as rider i has an individual willingness-to-walk distance $a_{ij} = a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i$, the overall conversion rate for block j during time window t is:

$$\begin{aligned}\mathbb{P}_{jt} &= E_\epsilon[\mathbb{P}_{ijt}] \\ &= \sum_{o=1}^{m_j} p_{oj}(\mathbf{W}) \cdot E_\epsilon[F_z(a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i | n_{jt}, b_{1oj}, b_{2oj})] \\ &\approx \sum_{o=1}^{m_j} p_{oj}(\mathbf{W}) \cdot \frac{1}{K} \sum_{k=1}^{K} F_z(a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \hat{\epsilon}_k | n_{jt}, b_{1oj}, b_{2oj}) \\ &\equiv G(a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}}, \sigma^2, \mathbf{W} | n_{jt}, b_{1oj}, b_{2oj}),\end{aligned} \tag{2.16}$$

where we approximate $\mathbb{P}_{jt}$ by numerically calculating the expectation of $F_z(a_{g(j)} + a_2 \cdot \mathbf{1}_{\{i \in L\}} + \epsilon_i | n_{jt}, b_{1oj}, b_{2oj})$. The approximation is performed to reduce the computation time for the optimal σ^2.

2.5.3 Estimation Results

We now construct our main structural estimation equation. Recall that in Equation 2.4, we divide customers into two types - those whose trips tend to be long-distance and those short-distance. We use γ_S to represent the proportion of potential riders who would initiate a short distance trip and $1 - \gamma_S$ the proportion that start a long distance trip. According to the Panel (b) of Figure 2.2, we observe a plummet in the number of trip after 1.2 kilometers. We thus classify trips within 1.2 kilometers as short distance trips and trips beyond 1.2 kilometers as long distance ones. Next, we use the rider type, the rider arrival rate from Equation (2.2) and the conversion rate derived from Equation (2.6) to construct expected number of transactions in each block-time pair. The underlying idea is to estimate the parameters $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \boldsymbol{\beta}, \mathbf{W}, a_U, a_M, a_D, a_2, \sigma^2, \gamma_S)$ by minimizing the sum of squared differences between the expected number of transactions predicted by our model and the number of trips documented in the data. Denote τ_{Sjt} as the number of short distance trips started in block j during time window t and τ_{Ljt} as the corresponding number of long distance trips, we write our estimation equation as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \sum_{j=1}^{J} \{(\lambda_{jt} \cdot \gamma_S \cdot G(a_{g(j)}, \sigma^2, \mathbf{W}|n_{jt}, b_{1oj}, b_{2oj}) - \tau_{Sjt})^2 + (\lambda_{jt} \cdot (1-\gamma_S) \cdot G(a_{g(j)} + a_2, \sigma^2, \mathbf{W}|n_{jt}, b_{1oj}, b_{2oj}) - \tau_{Ljt})^2\}, \tag{2.17}$$

where the summations are carried over all the time periods and blocks. We present the estimation results in Table 3.2.

Our estimation results show that for riders whose trip destinations are close, the average distances they are willing to walk at downtown, midtown and uptown areas to unlock a bike from Mobike are 25.751 meters, 28.99 meters, and 31.52 meters, respectively. In the meantime, our results show that $a_2 = 59.998$ and $\gamma_S = 0.893$, suggesting that 89.3% of the potential trips are short distance ones. The 11% riders who embark on long distance trips are will to walk for an additional 60 meters to unlock the bike, regardless of the location. The less-than-100-meter average willingness-to-walk highlights the importance of bike proximity to the riders. As for the demand density, we observe that each app-opening incidence is related to around three potential demand arrivals, which suggests that eyeballing is a more prominent bike-searching approach. Regarding to the potential rider origins, we see that most of the potential riders come from the residential areas ($w_{Res} = 1.919$), which conforms to the intuition that people commute from home in the morning rush hour. In contrast, commercial zones are inactive ($w_{Retail} = -0.054$) during the morning rush hour.

Table 2.4: Estimation Results

	Estimates	Std. Error
Demand Density		
Constant	1.597***	(0.233)
App-opening	2.977***	(0.060)
POI_Retail Store	-0.09***	(0.025)
POI_Restaurant	-0.011***	(0.001)
POI_Educational Institution	-0.002*	(0.001)
POI_Residence	0.005*	(0.002)
POI_Others	0.007	(0.013)
Hour_Dummy1	-1.282***	(0.262)
Hour_Dummy2	-0.863***	(0.211)
Weight_Retail Store	-0.054***	(0.019)
Weight_Restaurant	0.183***	(0.027)
Weight_Residence	1.919***	(0.495)
Weight_Others	0.012*	(0.006)
Consumer Utility		
a_U	31.52***	(2.665)
a_M	28.99***	(2.576)
a_D	25.751***	(2.78)
a_2	59.998***	(10.877)
σ	116.053***	(11.444)
γ_S	0.893***	(0.016)
Pseudo R^2	0.641	
Observations	45000	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Counterfactual Analysis

In this section, we first describe the framework we use for our counterfactual analysis in Section 2.6.1. Then, based on our estimation in Section 2.5, we conduct three counterfactual experiments in Section 2.6.2 - 2.6.4 to study the efficiency of the dockless bike sharing system in different market environments. The first counterfactual analysis examines the impact of the bike fleet size on rider conversion rate and system performance. We demonstrate that albeit larger fleet size generates more trips as potential riders will more likely to find a bike in their vicinity, deploying more bikes lowers the utilization rate for each bike and has a decreasing marginal return on ridership. More importantly, our framework would allow the Mobike to assess the impact of fleet size adjustment on customers conversion and ROI for each bike under various market conditions. In the second analysis, we propose a number of bike redistribution strategies, and demonstrate that the joint use of our structural analysis and our proposed bike redistribution strategy significantly increases the number of transactions over Mobike' current redistribution practice. In the last counterfactual, we compare the system performance of two common bike sharing models: the dock-based system and the dockless system. Our results highlight the pros and cons of both systems and identify the conditions that make one system more favorable than the other. Our results carry important implications for bikeshare companies as well as regulators to decide the most fitting bikeshare model for each city.

2.6.1 Setup

To track the movement of the bikes among different blocks, we first define the status of each activated bike. When a bike gets matched with a rider, it becomes unavailable to other potential riders until the trip is completed, after which the bike becomes a unit of supply at the destination block. Subsequently, we define the state of a bike at the beginning of a time period by a two-tuple (j, t), where j is the block that the bike will arrive after t number of time periods. For example, if a bike is in state $(1200, 0)$, this bike has arrived at block 1200 and is available at the current time period. Alternatively, if the bike is labeled as $(1200, 2)$, it indicates that the bike is currently in-transit (being used) and will be available at block 1200 in two periods. In what follows, we specify how we obtain the potential rider's origin, the destination of a trip and the associated timeline.

Rider Origin and Destination. According to our model, at the beginning of time period t, λ_{jt} potential riders show up in block j. For each potential rider, the specific square she spawns is randomly generated with probability $\{p_{oj}(\mathbf{W})\}$, and the exact origin coordinates are then uniformly generated within her origin square o. Notice that thanks to the "park anywhere" policy, the drop-off location of a transaction in the dockless system reveals the exact destination of the rider. Therefore, we obtain the empirical probability that a rider goes from block j to destination j' directly from our historical transaction data. Then, we draw the coordinates uniformly within the destination block as the rider's exact destination location.

Timeline. To be consistent with our estimation in Section 2.5, we discretize the morning rush hour (7:00 AM - 10:00 AM) into 10-minute time windows, and treat each 10-minute interval as the unit of time period in our analysis. As Figure 2.2 (a) shows that around 95% of the trips have duration shorter than 30 minutes, we assume that trips within the region of our analysis can only take up to 30 minutes. That is, if a bike gets matched at time window t, it would become available again no later than time window $t+3$, and can be available as early as $t+1$. We assume that trips are completed at the beginning of each period. Once a trip is completed, the bike will become available again at the destination block. Then, potential riders show up and make their commute decisions.

Bike Transition. Once a bike gets unlocked, its state will be governed by the trip duration and trip destination. When $t \geq 1$ at the beginning of the current period, the bike is already in a trip and still needs t periods until the trip completes. Then, at the beginning of the next period, the bike will need $t-1$ periods to arrive at the destination. Thus, we have:

$$P_{(j,t),(j',t')} = \begin{cases} 1 & (j',t') = (j,t-1), \\ 0 & \text{otherwise}, \end{cases}$$

where $P_{(j,t),(j',t')}$ denotes the transition probability from state (j,t) to state (j',t').

When the state $t=0$, it indicates that the bike is available at the beginning of the current period. Once matched, the bike's state at the next period depends on the trip duration and destination. To this end, we estimate the transition probabilities for states $(j,0)$ ($j \in \{1,\dots,J\}$) using the historical transaction data. Specifically, we construct the block-based bike movement data by recovering trip origin and destination blocks from the corresponding coordinates. Next, we classify the trips into three groups based on their trip duration. Group w ($w=1,2,3$) includes all the trips whose duration are within the interval $[(w-1)\times 10, w\times 10)$. For all the trips that cost more than half an hour, we classify them into group 3. Then, for state $(j,0)$, we collect all the trips that initiated from block j and calculate the empirical probability of transiting into each of the "destination block and trip duration group" combination.

Table 2.5: Example of Trip Information

Origin BlockID	Destination BlockID	Trip Duration Group	Counts
1	1	1	10
1	1	2	5
1	2	1	2
1	2	2	8
2	2	1	2
2	2	2	7
2	1	1	20
2	1	2	11

Example. We consider the aggregated between-block movements information as presented in

Table 2.5. According to Table 2.5, we have 4 states in total: $(1,0)$, $(1,1)$, $(2,0)$ and $(2,1)$. When an available bike gets activated, the empirical probability of transiting from state $(1,0)$ to state $(1,0)$ is $\hat{P}_{(1,0),(1,0)} = \frac{10}{10+5+2+8} = 0.4$. Here, we get the numerator 10 from the first row which represents the number of trips originated and ended both in block 1 within 10 minutes, and the denominator is the sum of all the trips that started from block 1. According to the same logic, we have the estimated transition probabilities as:

$$\hat{P}_{(1,0),(j,t)} = \begin{cases} 0.4 & (j,t) = (1,0), \\ 0.08 & (j,t) = (1,1), \\ 0.2 & (j,t) = (2,0), \\ 0.32 & (j,t) = (2,1). \end{cases}, \quad \hat{P}_{(2,0),(j,t)} = \begin{cases} 0.5 & (j,t) = (1,0), \\ 0.275 & (j,t) = (1,1), \\ 0.05 & (j,t) = (2,0), \\ 0.175 & (j,t) = (2,1). \end{cases}$$

Notably, there are several blocks that do not register any transactions during the span of our dataset, either due to lack of POI inside or the entire block being nature landscape where parking is forbidden. If somehow in our simulation a rider initiates a trip from such blocks, we assume that the trip is equally probable to go to the surrounding 9 blocks, and with the trip durations proportional to the average percentages in the rest of blocks (75% of the trips completed with 10 minutes, 18% of the trips completed within 10 - 20 minutes and 7% completed more than 20 minutes). Denote the total number of transactions started in block j at time window t as τ_{jt}, we summarize the procedures of our counterfactual analysis in Algorithm 3.

Algorithm 1:

input : Parameters: $\hat{a}_U$, $\hat{a}_M$, $\hat{a}_D$, $\hat{a}_2$, $\hat{\sigma}^2$, $\hat{\gamma}_S$, $\hat{\lambda}_{jt}$, $\hat{\mathbf{W}}$.

Initial snapshot of bikes, $\tau_{jt} = 0 \ (j = 1, \ldots, J; \ t = 1, \ldots, T)$.

Round the estimated $\hat{\lambda}_{jt}$ to the nearest integer as the number of arrivals.

for $t \in \{1, \ldots, T\}$ **do**

  for $j \in \{1, \ldots, J\}$ **do**

    Update the status of all the in-transit bikes;

    for $i \in \{1, \ldots, \mathit{round}(\hat{\lambda}_{jt})\}$ **do**

      Use transition probabilities $\hat{\boldsymbol{P}}$ to simulate the destination block j' and trip duration parameter w;

      Simulate i's origin and destination coordinates ;

      Find the nearest available bike in block j and calculate dis_{ijt};

      Simulate rider i's decision by drawing value a_{ijt} from the estimated distribution of willingness-to-walk distance, $Lognormal(\hat{a}_{g(j)} + \hat{a}_2 \cdot \mathbf{1}_{\{i \in L\}}, \hat{\sigma}^2)$;

      if $a_{ijt} \geq dis_{ijt}$,

      then Update the activated bike's status (w and coordinates), $\tau_{jt} = \tau_{jt} + 1$;

    end for

  end for

end for

2.6.2 Bike Fleet Size

In order to take over the market and win through the competition, dockless bike sharing companies tend to oversupply bikes in their operating regions (Ye [23], Brown [24]). However, the large influx of bikes raise concerns to the regulators as the piled bikes may block the pedestrians and jam the city (Taylor [25]), and a lot of cities have passed regulations to cap the bike fleet size (Jiang [26]). To better control the bike fleet and refine the bike deployment strategy, it is crucial to first understand how the system performance (e.g., customer conversion rate and bike utilization rate) gets affected by the number of deployed bikes across different areas.

In the first counterfactual analysis, we use the number of bikes and their distribution that come directly from our data as the benchmark. We vary the total number of bikes $\boldsymbol{N_0}$ in the initial distribution while keeping the proportion of bikes in each block the same to examine how

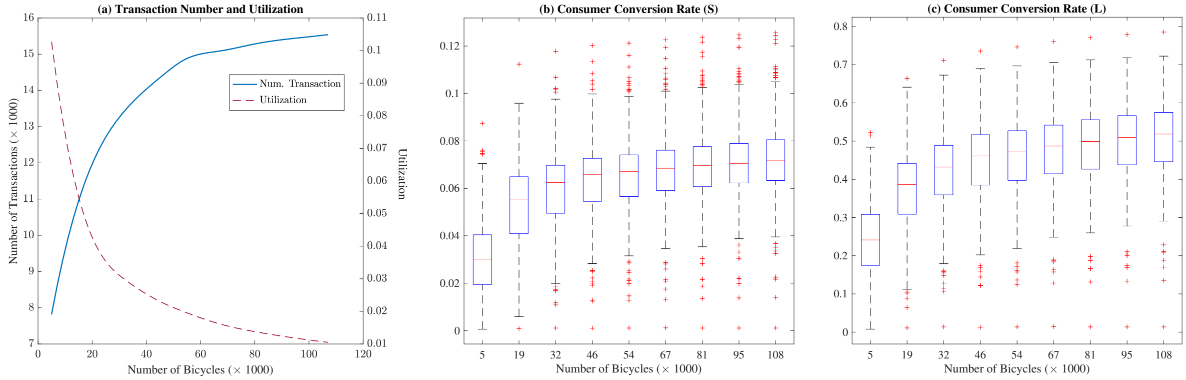

Figure 2.5: The impact of the number of bikes on rider conversion rates and the bike utilization.

the fleet size affects the system performance. Specifically, we focus on two performance metrics: the total number of converted transactions τ^{all} and bike utilization rate (UR), which is defined as:

$$UR = \frac{\sum_{t\in\{1,\ldots,T\}} \sum_{n\in\{1,\ldots,\boldsymbol{N_0}\}} \mathbb{I}_{\{\text{bike } l \text{ is in transit in period } t\}}}{T \times \boldsymbol{N_0}}, \tag{2.18}$$

where the indicator function $\mathbb{I}_{\{\text{bike } n \text{ is in transit in period } t\}} = 1$ when the bike n is used during t and zero otherwise, and $T \times \boldsymbol{N_0}$ is the total number of trips that $\boldsymbol{N_0}$ bikes can possibly supply over T periods.

We present the impact of the fleet size on rider conversation rates and the bike utilization in Figure 2.5. Panel (a) of Figure 2.5 suggests that the total number of transactions is increasing in the fleet size. Additionally, Panel (b) and (c) suggest that the block-level rider conversation rate (in both short and long distance groups) is increasing in the number of deployed bikes $\boldsymbol{N}_0$. This is because the walking distance for the potential riders to get to the closest bike decreases as more bikes are available in a block. Nevertheless, we also observe a diminishing marginal return from increasing the fleet size on both transaction volume and conversion rate. Meanwhile, the bike utilization rate decreases monotonically as the number of bikes surges, indicating that, despite the increase in conversion rate, deploying more bikes does not lead to a higher ROI per bike. Therefore, our model suggests that the company should consider reducing the current fleet size in the city of Changsha. Notably, when the fleet size is shrank by 40%, Mobike is still able to maintain 90% of its current transaction volume while saving up to $6.5 Million production costs in the city of Changsha according to Yang [27].

2.6.3 Bike Redistribution Strategy

The dockless bike sharing system provides riders the flexibility in the bike pick-up and drop-off processes. Nevertheless, such flexibility requires significant efforts on redistributing the bikes from the company. This is because without being constrained by docks, whose locations are carefully designed according to the number of POIs and numerous operational metrics, riders may drop off bikes at low-demand regions. To prevent bikes from constantly being idle or even piling up and becoming a menace to the city (Lazo [28], Brown [24]), the company needs to periodically trucking the bikes from niche areas to high-demand regions. Notably, Mobike has spent tens of million dollars on bike redistribution every month.

Specifically, the gist of any bike redistribution scheme is to identify move bikes from regions that have redundant bikes to areas that have deficit in supply. Thus, the performance of the redistribution strategy hinges heavily on the accuracy of demand forecast. In what follows, we first examine the effectiveness of Mobike's current bike redistribution strategy. We then study how our model framework which accounts for potential lost demand helps boost the effectiveness of the current bike relocation scheme.

We first describe Mobike's current relocation strategy. Let N_o denote the number of bikes at square o and $trans_o$ represent the average number of historical transactions the company observes at square o. The specific redistribution strategy is summarized as follows:

1. If $N_o - c_1 \cdot trans_o \geq m$, then square o is classified as an over-supplied square and $N_o -$

$c_1 \cdot trans_o$ bikes need to be moved away from square o.

2. If $N_o < c_2 \cdot trans_o$, then square o is classified as an under-supplied square and set $c_2 \cdot trans_o - N_o$ as square o's weight of getting redistributed bikes.

3. Relocate the over-supplied bikes to under-supplied squares in proportion to the weights we calculated in step 2 and place the transported bikes randomly within each corresponding square.

Here, c_1 and c_2 are two scaling factors deciding how much safety stock each square should keep, and m represents the minimum batch size that will trigger the redistribution for each square. A larger c_1 and smaller m imply that fewer squares and bikes are qualified as origin, while a larger c_2 implies that more squares should be included as the destination of bike relocation. These three parameters are set by the firm.

Since our structural model recovers the potential demand, the redistribution strategy we propose is identical to Mobike's current scheme, except that we replace the number of realized transactions in block o, $trans_o$, by the potential demand λ_o recovered from our model.

In this counterfactual experiment, we test the performance of the current transaction-based strategy (TB) as well as our demand estimation strategy (DE) using 27 combinations of $c_1,\ c_2$, and m. For each parameter combination, we follow either one of the TB and DE to relocate bikes. The bike distribution after the relocation is then used as the initial bike snapshot to conduct simulations. The results are presented in Table 2.6. As shown in the table, depending on the redistribution capacity, firm can adjust c_1 and m accordingly. Using our revised strategy, the better match between demand brought by more bike relocation can increase the number of converted trips by as much as 18.78% compared to the status quo. In contrast, the current strategy

Table 2.6: Performance Comparison between Different Redistribution Strategies

			m=0		m=5		m=10	
c_1	c_2	Policy	$\frac{\tau-\tau_{nr}}{\tau_{nr}}$	# of redist. bikes	$\frac{\tau-\tau_{nr}}{\tau_{nr}}$	# of redist. bikes	$\frac{\tau-\tau_{nr}}{\tau_{nr}}$	# of redist. bikes
Low=3	Low=1	TB	-19.4%	40706	2.07%	10253	1.94%	4224
		DE	14.89%	17830	9.19%	6037	5.45%	3219
	Medium=2	TB	-16.84%	40706	2.28%	10253	1.89%	4224
		DE	15.96%	17830	9.03%	6037	5.45%	3219
	High=3	TB	-16.08%	40706	2.3%	10253	1.67%	4224
		DE	18.78%	17830	9.99%	6037	5.97%	3219
Medium=5	Low=1	TB	-18.82%	38875	2%	9162	1.76%	3482
		DE	13.22%	11909	6.84%	4101	4.2%	2378
	Medium=2	TB	-16.38%	38875	2.13%	9162	1.61%	3482
		DE	13.73%	11909	6.61%	4101	4%	2378
	High=3	TB	-15.84%	38875	2.13%	9162	1.59%	3482
		DE	15.62%	11909	7.21%	4101	4.52%	2378
High=10	Low=1	TB	-18.64%	37534	1.91%	8170	1.51%	2807
		DE	8.95%	6474	3.91%	2165	2.39%	1292
	Medium=2	TB	-16.21%	37534	2.05%	8170	1.41%	2807
		DE	9.09%	6474	3.7%	2165	2.29%	1292
	High=3	TB	-15.84%	37534	1.92%	8170	1.29%	2807
		DE	10.01%	6474	4.26%	2165	2.61%	1292

Notes. τ represents the number of transactions for the corresponding bike redistribution strategy. τ_{nr} represents the number of transactions without bike redistribution. # of redist. bikes denotes the number of redistributed bikes.

boosts the system performance by only a thin margin when transporting bikes from squares that are flooded with idle bikes ($m = 5$ or 10), which is significant less effective compared to DE. Noticeably, it can even backfire and hurt the system performance when taking all the bikes away from the squares that have never generated any transaction ($m = 0$). The gap in efficiencies highlight the value of our model and the importance for the company to understand the potential demand.

2.6.4 Dockless vs. Docked Bikesharing

The dockless bike sharing system has proven to be an excellent last-mile urban solution in China thanks to its "park anywhere" policy. However, ridership for dockless bikes has been underwhelming in the U.S.. Although several companies such as Lime Bike adopt the dockless system, dock-based bike sharing still makes up to 80% of the bike sharing trips in the U.S. in 2019 (NACTO 2020 [2]). Compared to the dockless system, the docked model requires riders to pick up and return a bike at pre-installed stations. Capturing the key features of each bike sharing system, we investigate in this counterfactual experiment the efficiency of these two systems and highlight the managerial implications from our analysis.

2.6.4.1 Setup for the Dock-based System

We first specify the setup for the dock-based system. Compared to the dockless bike sharing system, the dock-based one by definition operates on a predetermined network of stations in the operating area. As a result, customers need to go to a non-empty station to pick up a bike and to a station with available docks to drop it.

Rider utility. In the dock-based system, the riders usually need to walk on the first and the last leg of the trip, i.e., from the rider's current location to the pick-up station and from the drop-off station to the actual destination. Thus, for a potential commuter i, who shows up in block j during time window t, the utility she would gain from using the sharing bike now becomes

$$U_{ijt} = a_{g(j)} + a_L \cdot \mathbf{1}_{\{h(i) \in L\}} - dis_{ijt} - dis^d_{ijt} + \epsilon_i, \tag{2.19}$$

where dis^{d}_{ijt} represents the walking distance of the second piece (from station to destination). Here, we assume that the disutility incurred by walking is linear to the total walking distance.

Station network. For the dock-based bike sharing system, the design of station network usually follows one of the two policies: the uniform rule and the demand-matching rule. According to He *et al.* [11], the city of London has been implementing the uniform rule, where any two adjacent stations are at most 300 meters away from each other across the city. While in New York city, the Citi bike network has higher density in the downtown area. The system covers only the downtown and midtown areas in Manhattan, and very few places in uptown Manhattan and Brooklyn. For this counterfactual experiment, we consider both station network design policies. We set the number of stations $\mathbf{K} = 10,000$ in our analysis. Then, on average, every 25 squares will contain one station.

Given the number of stations $\mathbf{K}$, we first evenly spread the stations across the operating region. This means that the adjacent stations will be 250 meters away from each other. We also evenly assign the $\mathbf{N_0}$ bikes to all the stations as the initial distribution. We determine the number of docks in each station based on its initial inventory of bikes and make sure that the dock-to-bike ratio is around 1.5-to-1 (Shaheen *et al.* [29]). We refer to this network and bike initial distribution setup as the uniform network ("UN").

Next, we design and examine the effectiveness of a separate station network based on the demand-matching rule. Specifically, we first assign stations to all the blocks in proportion to the potential demand that we estimated in Section 5.3. Then for a block that has k stations, we rank the squares by their demand densities, select the first k squares, and put one station at the center of each of them. We then assign the $\mathbf{N_0}$ bikes to all the stations according to the demand density of the squares. Same as in UN, we set the number of docks in each station as 1.5 times its initial inventory of bikes. We refer to this one as the demand-matching network ("DN").

Dock availability. Dock availability is another constraint that we include in the dock-based system. When a rider is about to complete her ride, she will first consider returning the bike back to the nearest station from her destination. However, the dock availability of that specific station is uncertain. When the docks are all full, she has to compromise and drop off the bike in another station that has available dock. We estimate the dock availability by the following procedure. We first assume that all stations in the same block share the same dock availability and then use the historical bike snapshot data to calculate the number of bikes in each block at all snapshots. Whenever the number of bikes exceed the capacity (station size $\times$ number of stations in block j) of block j, we consider all the stations in block j as dock-unavailable, otherwise they are all available. In this way, we estimate the dock availability of each station as

$$dock\hat{\_}avail_j = \frac{\sum \mathbb{I}\{n_{jt} \leq \text{Station size} \times \text{Number of stations in block } j\}}{\text{Number of snapshots}}. \tag{2.20}$$

Riders' decision-making process. When a rider shows up in the dock-based system and tries to make commuting decision, in order to calculate the utility that she can get from using the system, she needs the two pieces of walking distance (dis_{ijt} and dis^d_{ijt}). For dis_{ijt}, it would be the Manhattan distance between the customer and her nearest non-empty station at period t. However, because of the possibility that a station may be full of bikes when the customer completes the trip, it is not determined which station the customer can return the bike back to.

Therefore, to calculate the expected walking distance after the drop-off, we use:

$$
\begin{aligned}
dis^d_{ijt} = &\sum_{s=1}^{S_j} d_{si} \times dock\_avail_{sj} \times \Pi_{u=0}^{s-1}(1 - dock\_avail_{uj}) \\
&+ M \times \Pi_{u=0}^{S_j}(1 - dock\_avail_{uj}),
\end{aligned} \tag{2.21}
$$

where S_j represents the station index set for block j, d_{si} is the walking distance from station s to rider i's destination, and M is the average walking distance for customer i when there is no dock-available station in j, which we set to 250 meters.

Simulation. We now describe how we run the simulation for the dock-based bike sharing system. In the dock-based system, since the station network is given, in order to monitor riders' decision process and bike movements, it is important to keep track of the number of bikes at each station.

Similar to the setup in Section 2.6.1, we use τ_{jst} to denote the total number of transactions started from station s of block j at time window t, ν_{jstw} to represent the number of transactions that start at period t and finish in station s of block j at time window $t + w$ ($w = \{1, 2, 3\}$). In the meantime, we use I_{jst} to represent the number of bikes that will arrive in station s of block j at period t. The number of arriving bikes in period t come from three sources: 1) the newly activated bikes from period $t - 1$ whose transition time is one period, $\nu_{j,s,t-1,1}$, 2) the activated bikes from period $t - 2$ whose transition time are two periods, $\nu_{j,s,t-2,2}$, and 3) the activated bikes from period $t - 3$ whose transition time are three periods, $\nu_{j,s,t-3,3}$. To sum, we have

$$
I_{jst} = \sum_{w=1}^{3} \nu_{j,s,t-w,w}. \tag{2.22}
$$

We use $\boldsymbol{N_{jt}} = (n_{j1t}, n_{j2t}, \ldots, n_{jS_jt})$ to denote the number of available bikes at each station

Algorithm 2:

input : Parameters: $\hat{a}_U$, $\hat{a}_M$, $\hat{a}_D$, $\hat{a}_2$, $\hat{\sigma}^2$, $\hat{\gamma}_S$, $\hat{\lambda}_{jt}$, $\hat{\mathbf{W}}$.
Initial number of bikes $\mathbf{N_{j0}}$, $\tau_{js0} = 0$, $I_{js0} = 0$ and ν_{js0w} for $w = 1, 2, 3$.
Round the estimated $\hat{\lambda}_{jt}$ to the nearest integer as the number of arrivals.

for $t \in \{1, \ldots, T\}$ **do**
  for $j \in \{1, \ldots, J\}$ **do**
    for $s \in \{1, \ldots, S_j\}$ **do**
      Update the number of arriving bikes: $I_{jst} = \sum_{w=1}^{3} \nu_{j,s,t-w,w}$;
      Update the available number of bikes: $n_{jst} = n_{js,t-1} - \tau_{js,t-1} + I_{jst}$;
    end for
    for $i \in \{1, \ldots, \mathit{round}(\hat{\lambda}_{jt})\}$ **do**
      Use transition probabilites $\hat{\boldsymbol{P}}$ to simulate the destination block j' and trip duration parameter w;
      Simulate i's origin and destination coordinates ;
      Find the nearest available station s and calculate dis_{ijt};
      Calculate dis^d_{ijt} and get returning station s' by using Equation (2.21);
      Simulate rider i's decision by drawing value a_{ijt} from the estimated distribution of willingness-to-walk, $Lognormal(\hat{a}_{g(j)} + \hat{a}_2 \cdot \mathbf{1}_{\{i \in L\}}, \hat{\sigma}^2)$;
      if $a_{ijt} \geq dis_{ijt} + dis^d_{ijt}$,
      then Update $\tau_{jst} = \tau_{jst} + 1$, and $\nu_{j's'tw} = \nu_{j's'tw} + 1$;
    end for
  end for
end for

of block j at the beginning of time window t, where S_j represents the number of stations in block j. In period t, τ_{jst} bikes are picked up from station s. Then in $t+1$, $I_{j,s,t+1}$ activated bikes will arrive at station s of block j and become available again. Therefore, to update $n_{j,s,t+1}$, we have:

$$n_{j,s,t+1} = n_{jst} - \tau_{jst} + I_{j,s,t+1}, \tag{2.23}$$

and the we describe the procedure of our counterfactual analysis in Algorithm 2.

2.6.4.2 Monopoly Setup

To assess how different bike sharing rules affect the system performance, we run our counterfactual in the monopoly market setup. Notice that we observe the data in a competitive market where there exists more than one dockless bike sharing company. Thus, the estimation in Section 2.5 only identifies the willingness-to-walk to bikes from Mobike. In order to reveal riders' willingness-to-walk to any sharing bike, we need to consider all the sharing bikes in the market.

To this end, we consider bikes from different bikesharing companies identical. Also, we make the assumption that all bikes have the same distribution as Mobike across blocks and squares. Following this, the total number of bikes in block j at time t is $n_{jt}^A = n_{jt}/share_m$, where the superscript 'A' denotes variables related to all the bikes, and $share_m$ represents the percentage of bikes from Mobike. In our specific case, we assume $share_m = 50\%$. In practice, companies can refine its estimation using more accurate market data. Importantly, since all bikes have the same distribution as Mobike, they also have the same bike-to-rider distance distribution as Mobike. Therefore, the overall conversion rate for Mobike at block j during time window t becomes:

$$\begin{aligned}\mathbb{P}_{jt} &= share_m \cdot \mathbb{P}_{jt}^A \\ &= share_m \cdot G(a_U^A, a_M^A, a_D^A, a_L^A, {\sigma^A}^2, \mathbf{W} | n_{jt}^A, b_{1oj}, b_{2oj}).\end{aligned}$$

Then, we can use the following estimation equation to estimate the willingness-to-walk to any

Table 2.7: Estimation Results

	Estimates	Std. Error
Consumer Utility		
a_U^A	70.879***	(4.016)
a_M^A	64.828***	(3.783)
a_D^A	56.128***	(3.965)
a_L^A	194.19***	(14.84)
σ^A	193.67***	(9.524)
Pseudo R^2	0.633	
Observations	45000	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

sharing bike:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T}\sum_{j=1}^{J}\{(\lambda_{jt} \cdot \gamma_S \cdot share_m \cdot G(a_U^A, a_M^A, a_D^A, a_L^A, {\sigma^A}^2, \mathbf{W}|n_{jt}^A, b_{1oj}, b_{2oj}) - \tau_{Sjt})^2 + (\lambda_{jt} \cdot (1-\gamma_S) \cdot share_m \cdot G(a_U^A, a_M^A, a_D^A, a_L^A, {\sigma^A}^2, \mathbf{W}|n_{jt}^A, b_{1oj}, b_{2oj}) - \tau_{Ljt})^2\}.$$

The results are presented in Table 3.3.

2.6.4.3 Results

To examine how the "park anywhere" policy in the dockless system and the station network in the dock-based system affect the bike movements, shape the bike distribution and hence affect the system performance differently in the long run, we consider a prolonged time span and let the system evolves on its own (without any intervention, e.g., bike redistribution). In order to make sure everything is the same except the system setup, we also keep the initial bike distribution in the dockless system the same as in the corresponding dock-based setup. Then, for three

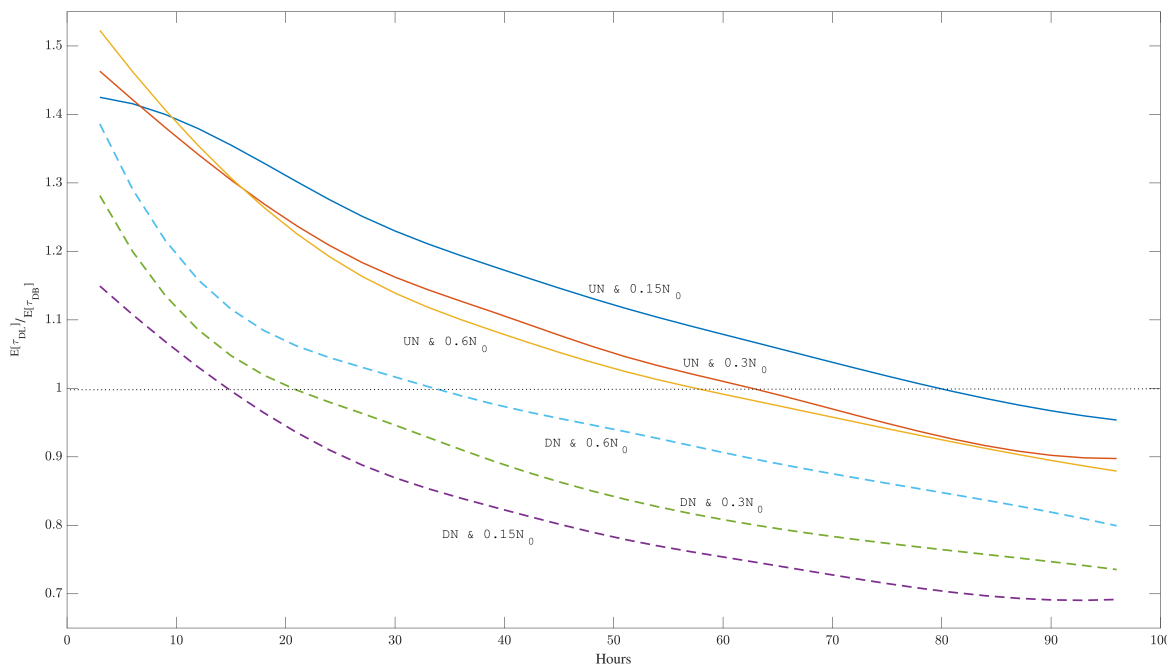

Figure 2.6: Transaction Performance Ratio between Dockless and Dock-based Systems

different fleet sizes ($0.6\boldsymbol{N_0}$, $0.3\boldsymbol{N_0}$, and $0.15\boldsymbol{N_0}$), we conduct the counterfactual and calculate the performance ratio (the ratio between the number of transactions in the dockless system and its corresponding dock-based system). The results are presented in Figure 2.6.

As can be seen in Figure 2.6, regardless of the bike fleet size and the dock station location distribution, the dockless system has a substantial advantage over the dock-based system early on: the converted number of transactions in the dockless system can be over 50% higher compared to the dock-based system. This is consistent with the intuition that the dockless system is more appealing to riders. Its edge comes from the "park anywhere" policy, which eliminates the additional walking (from station to destination), which in turn incentivizes more riders to use bikes even if they are farther away from available bicycles.

However, this convenience for riders comes with the cost of having less control of the bike distribution. Since each bike's location now evolves based on previous riders' actual destinations, its odds of being picked up and used again is highly dependent on where it was left at. In the extreme case, the bike will be stuck an stay unused for a long time (even indefinitely) once it is dropped off at a place where very little bike demand originates. As long as there exists some low-demand trip destinations, as time goes on, more and more bikes can get stuck and pile up in an effective bicycle "black hole" and thus the system efficiency will be hurt. This is why the transaction ratio between the dockless and the dock-based systems decreases with time. It also implies that a frequent system maintenance (e.g., bike redistribution) is required for the dockless system to sustain a high service level relative to a dock-based system.

In contrast, in the dock-based system, bike drop-off locations are constrained by the locations of the docking stations in the network. At first glance, this setup is discouraging to the potential riders since it takes the flexibility away from them. Nevertheless, the station network disciplines the bike drop-off behavior and maintains a certain level of the bike accessibility for the system. As illustrated in the figure, the performance level of the dock-based system drops in a slower rate than the dockless system. The performance gap between the two systems is narrowing. In particular, when the network of stations is carefully designed, i.e., matching the potential estimated demand, the performance of the dock-based system eventually overtakes the dockless one. To put it in another way, squeezing riders' IR (individual rationality) constraints for bike-share usage by forcing them to return the bikes in predetermined locations (docks) actually provides them a benefit in the long run, in that bikes become more easily available to them and the system becomes more attractive on average, leading to a higher number of transactions. This advantage is more influential when the bicycle fleet size is limited, where temporary bicycle movements will determine a big chunk of the future supply allocation.

In order to demonstrate how docking station location choice alleviates the bicycle imbal-

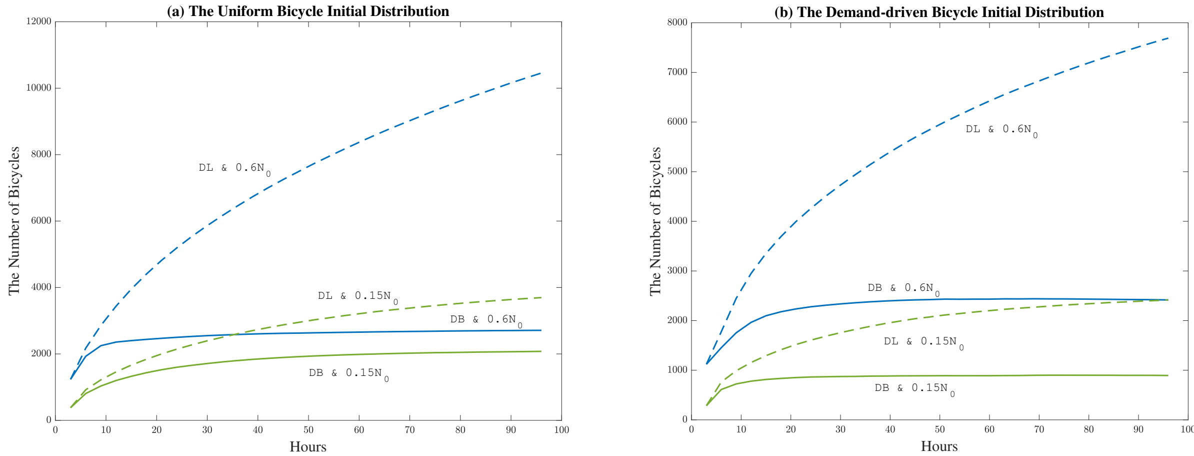

Figure 2.7: The Number of Bicycles in the Top 5% Low-Demand Blocks

ance issue, we monitor two metrics that measure the bicycle pileups and the overall imbalance between bicycle supply and demand. First, we can study the bicycle pile ups that are known to significantly hinder the dockless system performance. In order to do this, we define a *low-demand block* as a block that has lower-than-average trip origination demand level in the time window studied. Keeping track of the total number of bicycles in the top 5% of the low-demand-blocks that have the highest number of bicycles, we can observe the formation of bicycle pileups, which is demonstrated in in Figure 2.7.

As can be seen in the figure, bicycles pile up much more rapidly in the dockless system compared to the dock-based system for both uniform and demand-driven bicycle distribution cases. What is more, having more bicycles in the system does not help and in fact makes things multiple times worse. Two things to notice are that, first, with the demand-driven bicycle distribution strategy, the pileups occur much slower than the uniform bicycle distribution strategy, which demonstrates significant gains from the smarter system design. Second, the dockless system is much better in containing pile ups than the dockless system. This is partly due to the limited number of docks in the lower demand areas, which in the long run caps the number of bicycles that can be parked in such parts of the city. However, as can be seen from the figure, even in the short run, the dock-based system is better in controlling undesired pilups, because it is better at controlling the rider behavior since riders would have less incentives to take bicycles to low-demand areas if they know they will have to find a dock far away from their destination and walk a significant amount if they do so.

We can expand the perspective and next take a look at the bigger picture by studying the imbalance level of the whole system. To do this let us define the *Bike Imbalance Index* (BII), as

$$\mathrm{BII}_i = \sum_{j=1,\ldots,J} (\lambda_{ij} - N_{ij})^2, \tag{2.24}$$

where λ_{ij} and N_{ij} represent the expected bicycle demand and current bicycle supply for block j in time segment t. BII in a sense provides a measure of squared deviation from the ideal bicycle distribution based on expected demand, i.e., higher BII indicates a more imbalanced system with more underserved or overserved blocks or both. We monitor the evolution of BII for different scenarios to see the effects of bicycle numbers system design. Figure 2.8 presents the results.

As can be observed from Figure 2.8, with the exception of early periods for the uniform bicycle distribution case, the dock-based system always has a smaller BII compared to the dockless system. This is because, as we discussed above, the dock-based system is in general better in keeping the bicycles in higher demand locations, albeit at the expense of transactions. However, as can be seen in panel (a), when the bicycles are uniformly distributed, the BII for the dockless system initially rapidly *decreases*. This is because the dockless system stimulates transactions, which initially move bicycles from lower demand to higher demand locations since in general

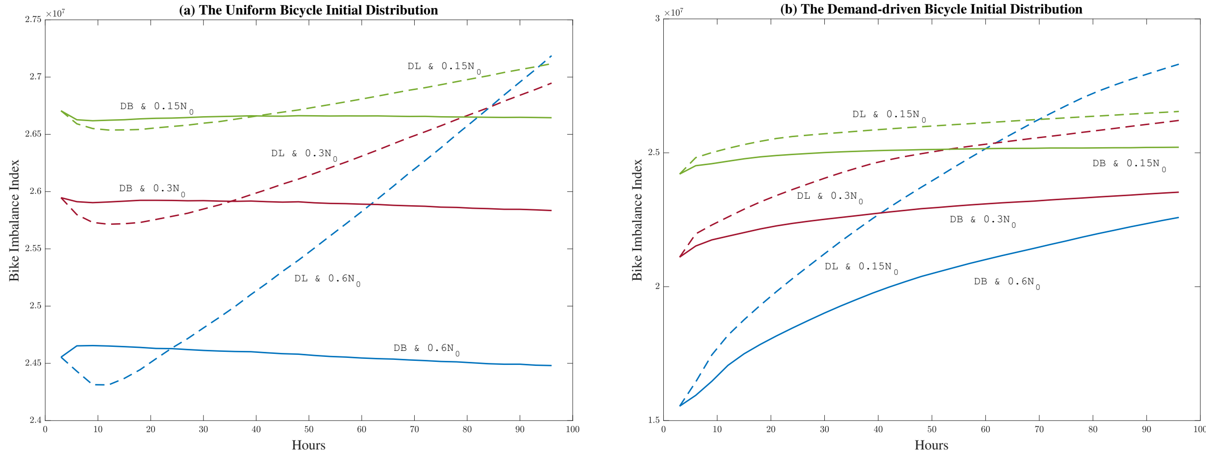

Figure 2.8: Bike Imbalance Index Comparisons

higher demand origination locations are also popular customer destinations. Note that this BII decrease also happens with the dock-based system but at a much slower pace. As a result, the BII for the dockless system becomes better (lower) for the dockless system compared to the dock-based system. However, as time progresses, the tendency of the dockless system to spread the bicycles towards lower demand areas dominates and dockless system BII becomes sigificantly wrse than that of the dock-based system. Comparing panels (a) and (b) of Figure 2.8, an important thing to observe is how much better-balanced and more stable the system becomes when the initial bicycle distribution is demand driven. In panel (b) we can see that for all fleet sizes, the demand driven system not only has a better BII, but for both dockless and dock-based systems, its BII is more stable over time. This shows that a smarter initial bicycle and dock location distribution can have significant and lasting postive effect on system performance.

To summarize, the "park anywhere" policy is appealing to riders because it injects the flexibility into the bike sharing system. It boosts up the converted trips when having the same bike distribution. However, it makes the dockless system more vulnerable to low-demand trip destinations and the fleet size in terms of long term performance. Meanwhile, the dock-based system has an edge on disciplining rider behaviors by having a network of stations, which imposes better control on the bike distribution of the system. Therefore, all else equal, the dockless system is able to achieve better performance than its docked counterpart, yet it requires constant bike redistribution to remain efficient in the long run. The dock-based system, however, can be the preferable one when the operational intervention (e.g., bike redistribution cost) is too costly and the bike fleet size is moderate.

2.7 Conclusions

We study Mobike's dockless bike sharing system in Changsha. Different from the dock-based system which has a pre-installed network of stations, bikes can be parked anywhere in the operating region for the dockless system. For this reason, the state space of each bike is infinite and it is infeasible to adopt the classic demand estimation framework developed for the dock-based system. To tackle this challenge, we endogenize the distribution of the bike-to-rider distance and develop a utility-based empirical framework to model the rider decision process. The method can be extended to studying demand for other dockless micromobility systems. We then empirically estimate customer's willingness-to-walk to pick up the bike and recover the potential customer demand.

Using our model and the estimated parameters, we provide practical guidance on the bike fleet size of the system. Our analysis suggests that the system performance level is increasing in the bike fleet size in a diminishing marginal rate. In particular, the company can still maintain 85% of its current service volume while saving millions of production costs and other operational costs if reducing the fleet size by 40% in Changsha. Our empirical results provide important insights and recommendations to the bikeshare companies and regulators.

To prevent bikes from constantly being idle or even piling up and becoming a mess to the city, dockless bike sharing systems put a lot of effort on bike redistribution. We evaluate a (S, s) policy-like strategy of the bike redistribution for Mobike. We find that the current strategy, which incorporates historical transaction information, boosts the system performance by only a thin margin and sometimes can even hurt the system performance. Instead, we propose a revised strategy to integrate our estimated potential demand and show that it can increase the number of converted trips by as much as 45.69% compared to the status quo. Our results demonstrate the value of understanding the potential customer demand, especially in making operational decisions.

We finally shed some light on the comparison between the dockless bike sharing system and the dock-based system. Capturing the key features of each bike sharing system, we highlight the tradeoff between parking flexibility and disciplining bike distribution. We show that the "park anywhere" policy provides the dockless system with an advantage on attracting riders over its docked counterpart. Compare with the dock-based system, riders are willing to walk longer distance to pick up a bike in the dockless system. It makes the dockless system much more efficient than the dock-based one in the short run. At the same time, though, the "park anywhere" policy gives away the control of bike accessibility and makes the dockless system more vulnerable to trip destinations.

On the other hand, the pre-installed network of stations helps the system to discipline the potential locations of the bikes. Although it sacrifices part of the utility from the rider end, the dock-based system obtains an edge on achieving better match between supply and demand in the long run. Our counterfactual results reveal that given the same initial condition, the dockless system starts with a much better performance than the dock-based system. As time goes on, more and more bikes in the dockless system are absorbed into the ‘black hole’ areas (trip destinations that have few demand) and hence the performance gap narrows. After some point, the dock-based system becomes the better performed system. Therefore, to sustain its high service level, the dockless system needs regularly bike redistributions. Larger bike fleet size can also prevent its service level from falling too fast. Our analysis carries important implications for bikeshare companies as well as regulators to decide the most fitting bikeshare model for each city.

Our work also suggests several issues that require further study. First, while our analysis emphasizes the morning rush hour in the daily commute, it would also be valuable to include other time slots of the day and evaluate different preference patterns during the day. Second, our comparison between the two bike sharing systems assumes no redistribution work or any other interference. Including the bike redistribution work in the comparison would be more helpful in understanding the cost each system needs to maintain a certain system performance level, hence providing more insights on choosing the right bikeshare business model under various scenarios. Finally, the coexistence of both types of the bike sharing system and the hybrid model should be studied. We hope that our work will stimulate research on these interesting and important topics.

Chapter 3: Improving Profit through Improving System Performance in Bike Sharing

3.1 Introduction

With urbanization on the rise, the majority of trips people take fall within the category of micromobility and thus are prime candidates for bike and scooter usage. And as consumers take advantage of this growing trend, the market opportunity continues to expand. In the US alone, customers took 36.5 million trips on dock-based bike sharing systems in year 2018 (NACTO.org). In 2019, 109 cities had dockless scooter programs, a 45% increase from 2018, which contributed to an over 100% increase in trips taken on scooters nationwide. As the e-scooter trend continues, the micromobility market in the US is predicted to be worth between $200-$300 billion by 2030, per McKinsey.

Asia has been a leading pioneer in the micromobility world, with China being the first country to implement a dockless bike-sharing platform in 2015. With less regulatory red tape in place compared to Europe and North America, micromobility startups had the advantage of quick implementation across cities in Asia. This lack of regulation, however, also resulted in an oversaturated market, with millions of bicycles piling up on city streets. By the end of 2019, dockless bike sharing systems have served more than 360 Chinese cities with 47 million daily trips (State Information Center of China 2019). Since 2018, bike-sharing companies in China began shifting their focus to e-bikes, rather than regular bicycles, and set up charging kiosks across their operating areas.

Though many micromobility companies raked in millions of dollars through investors, many struggle to achieve sustainable profitability. Ofo is still buried under cash flow problems. Although its app has since pivoted to become an e-commerce platform, the company owes refunds to millions of users who placed deposits on rides. Some startups believe that a subscription-based business model might be a path to profitability, as subscriptions can be a more dependable recurring revenue stream at a potentially higher price point. Aiming to maintain a constant supply of bicycles in the right places, improve user experience, and hence retain more subscribers, bike-sharing companies often spend a lot of money and effort to flood the market with an increasing number of bicycles. According to the Ministry of Transport of China (MOT), almost 20 million sharing bicycles have been deployed into the domestic market till the end of October 2020. The number was once growing so fast that some local governments had to step in and provide guidance to make sure it was within city capacity.

Meanwhile, bikeshare firms also manually redistribute bikes across different regions. In fact, the bike relocation scheme has been widely used as a daily operational tool in bike sharing systems. However, although high bike accessibility potentially leads to more converted trips, the effectiveness of the redistribution work (better system performance) on revenue growth is unclear. First, the redistribution scheme is costly. For instance, Meituan Bike spends tens of millions of dollars on bike redistribution every month. Therefore, it requires a cost-efficient scheme to improve the system service level. Second, the relationship between the service level and subscription rate is not clear. If most of the customers choose to purchase or renew their subscriptions as long as the service level reaches a threshold that can be maintained without redistribution, then the intensive rebalancing work does not make much economic sense.

To answer these questions, we study the causal relationship between revenue and performance of the bike sharing system, and aim to find the bike fleet size and subscription price that maximize the total revenue. In this paper, we first identify the effect of system performance on customer subscriptions, specifically the probability of customers purchasing different subscription plans. Then building on the structural framework developed in Chapter 2, we study how the system service level and subscription price affect customers sign-up decisions. The primary goal of this paper is to empirically measure the economic value of system performance and utilize customer preferences in the operational decision-making process.

3.2 Literature Review

Our paper is related to the emerging literature stream on bike sharing, an important topic of sharing economy. Bike-sharing has received increasing attention, especially from the transportation area, in recent years with initiatives to increase cycle usage, improve the last-mile connection to other modes of transit, and lessen the environmental impacts of our transport activities. DeMaio [1] and Shaheen *et al.* [30] discuss the history of bike-share systems and its current trends. Dell'Olio *et al.* [31] proposed a methodology for calculating the potential demand for bicycle use and other related factors before implementing bike-share systems. With a similar goal but different approach, Daddio [32] studies the determinants that will affect the usage rate empirically. Gebhart and Noland [33] look at the specific implact of weather conditions. Both papers use the data from Capital Bikeshare, one of the largest bike-share porgram in the US. For surveys of this literature, readers are referred to Fishman *et al.* [34] and Fishman [35]. This literature usually uses qualitative analysis and focuses more on the customer preferences and behaviors, safety concerns, and some other transportation-related topics. However, our work studies how operational excellence translates into higher profit in the bikesharing industry and is thus related to the burgeoning literature in bike sharing. Bikesharing companies usually operate in one of the two typical models: docked-based or dockless. In the docked-based system, bikes are parked at designated stations. To better match supply with demand in such a system, Chemla *et al.* [16], Dell'Amico *et al.* [17], O'Mahony *et al.* [18], Schuijbroek *et al.* [19] and Ghosh *et al.* [20] study how to relocate bikes among stations. The relocation can be either one-shot or continuous over time. Meanwhile, Kabra *et al.* [10] and He *et al.* [11] study the demand estimation problem in docked-base bikesharing system by treating each bike station or each orgin-destination station pair as a distinct product. Different from the docked-base system, bikes can be parked freely in a dockless system. In this case, each bike station can no longer be treated as a product as there are almost infinite ”stations” for parking.

The other main issue examined in this paper is subscription. This paper contributes to the growing literature in how bikeshare firms should utilize different operational levers to achieve maximized and sustainable profit. Our work studies the relationship between the bike service level and profit through the channel of the number of subscribing users. Thus, our work is related to the vast operations and marketing literature on the effect of pricing and payment methods on subscriptions. Pattabhiramaiah *et al.* [36] demonstrate that offering free access to online pay-walled news content improves the retention of newspaper subscribers and boosts the company's revenue. Danaher *et al.* [37], Tian and Feinberg [38], Birge *et al.* [39], Kash *et al.* [40], and Mai and Hu [41] study how pricing and discount affect subscription rate. In addition to the effect of price, we investigate how operational performance, in our case the service level of bikes, affects user retention and new sign-ups. Our goal is to maximize the profit by balancing the cost of maintaining a certain degree of service level and the revenue from the subscription level which is in turn affected by the service level.

3.3 Empirical Setting and Data

3.3.1 Empirical Setting: The Meituan Bike System

To study how the system performance and price affect subscription behavior, we use data from Meituan Bike. Meituan Bike (previously known as Mobike) is one of the largest bike-sharing companies in the world. Registered riders on Meituan Bike search for and unlock available bicycles using the Meituan Bike APP when they need a bike ride. Like most bike-sharing services, Meituan Bike's current business model operates with different ride options. It not only provides the single-ride option, for which riders pay per ride based on the trip duration, but also offers multiple membership plans (monthly pass, seasonal pass, annual pass, etc). Take the monthly pass (which is also the most purchased plan) as an example. A rider will need to pay a subscription fee beforehand to become a subscriber, and then gets unlimited free rides in the following 30 days. The subscription/membership plans are an important approach to prevent frequent riders from leaving for other competitors and generate consistent revenue streams. As of 2021, there are more than one million subscribers nationwide in China and 10 million riders who use the service.

As a result of the intense competition and the desire to achieve sustainable profitability, it becomes a major concern, for bike-sharing companies, to choose a price point and maintain a delicate level of system performance (e.g., decisions on bicycle fleet size and bicycle redistribution schedules) so that the profit is maximized. The main focus of our study is to use data from Meituan Bike as an example to study the effects of system performance and price on subscription behavior that can be utilized in optimizing operational decisions.

3.3.2 Meituan Bike Data: Subscription Behavior and Service Level

To conduct this study, we collect data from a randomly selected sample of 19,934 Meituan Bike registered users that had ever used the service during the timespan (January 1st, 2021 to August 31st, 2021) in Beijing. For each one of these users, we monitor their APP-related activities throughout the eight-month period and construct the sign-up events that will be used in our estimations. We restrict our attention to the monthly subscription plan and exclude other plans with different durations and pricing patterns. This is because the monthly pass dominates among all various plans and contributes a big chunk of the revenue, and other plans tend to attract riders with different needs and preferences.

To analyze bike users' subscription purchasing behavior, we collect sign-up events (potential purchasing events) from each rider's historical activities. We classify the subscription behaviors into two types: 1) *subscription renewal events*, the decisions from the existing subscribers who already purchased the monthly pass and are still within the effective period, and 2) *new sign-up events*, the joining decisions from the registered non-subscribers who may or may not have purchased the monthly pass before but currently are not subscribers and face the joining or not decisions. In this paper, we define an existing subscriber as one who still possesses an unexpired monthly pass and consider all other riders as non-subscribers. Therefore, if a rider has

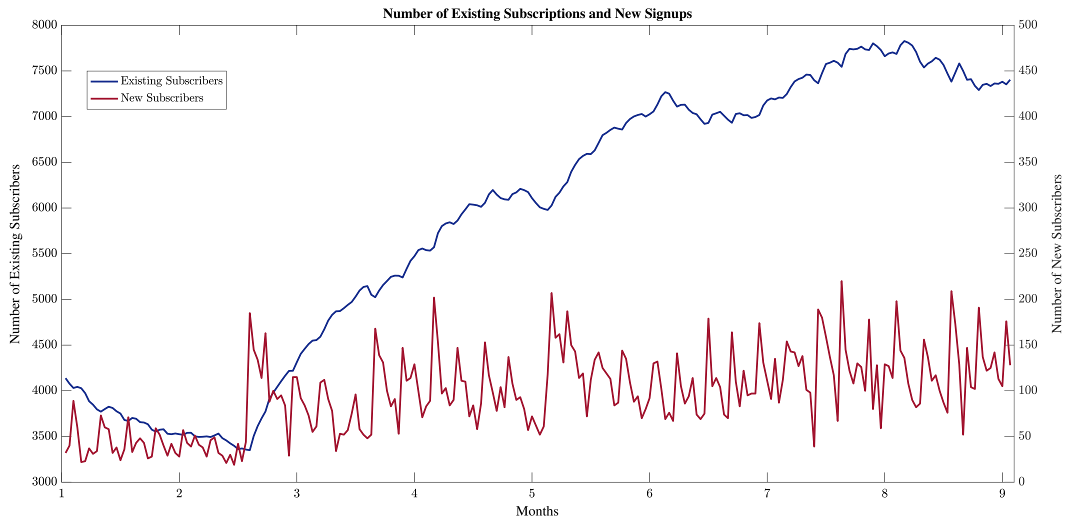

Figure 3.1: Number of Existing Subscribers and New Sign-ups

a monthly pass that is going to expire on day $t + 1$ and decides not to renew before day t, we consider him/her as an existing subscriber on day t and a non-subscriber on day $t + 1$. Figure 3.1 presents the evolution of both the number of existing subscribers and new sign-ups throughout the eight-month period.

3.3.3 Subscription Behavior

The total number of subscribers consists of the unexpired existing subscribers and new sign-ups. To keep track of the subscribers, we need to understand the subscription behavior from both the existing subscriber renewals and the new sign-ups of the non-subscribers.

An existing monthly subscriber can extend (purchase another monthly pass) her subscription anytime within the 30-day time window. The newly purchased membership will begin right after the current one ends. Therefore, as long as it is before the expiration day of the current subscription, we consider this membership extension as a subscription renewal. On the other hand, if no renewal happens before the subscription expires, the rider becomes a non-subscriber and we

Table 3.1: Summary Statistics

Subscription Events	Mean	St. Dev.	Min	Max
Price (CNY per day)	0.49	0.12	0.00	0.83
Service Level (sl)	0.49	0.17	0.00	1.00
Number of Trips	0.49	1.08	0.00	29.00
Number of APP-related Activity	0.94	2.04	0.00	83.00

consider the rider opting out of the plan. Following this logic, we monitor the subscription status of all the riders and construct the sign-up events for all existing subscribers.

The sign-up events for non-subscribers are trickier to define. Note that, unlike the existing subscribers who always need to make the renew-or-not decision at the expiration day, the sign-up events for non-subscribers can not be observed unless the rider chooses to opt in. Therefore, on each day t, we consider that all the active (the formal definition will be introduced later) non-subscribers face the purchase/no purchase decision. Riders who decide not to become a member are the ones who are still non-subscribers on day $t+1$.

3.3.4 Service Level

Riders' purchasing decisions can be affected by a lot of factors: subscription fee, the unit price per ride, his/her needs for rides during the potential subscription period, etc. The factors we mainly focus on are the rider's perception of service level, which we define as

$$sl = \frac{\text{\# of converted trips in the last 30 days}}{\text{\# of bike searching behavior in the last 30 days}}, \tag{3.1}$$

and subscription fee. sl measures the bike accessibility of each individual user for the last 30 days. In this study, we use the number of APP opening behavior (which is recorded and collected by the company) as the proxy of the bike searching behavior. Table 3.1 displays the summary statistics.

3.4 Econometric Specifications and Results

Fleet size control and bike relocation schemes have been widely used as operational tools in bike-sharing systems. Companies spend a lot of money and effort on deploying/maintaining more bicycles in the market network and redistribution work, trying to improve the system service level. However, the relationship between the service level and subscription rate is not clear. If most of the customers choose to purchase or renew their subscriptions as long as the service level reaches a threshold that can be maintained without having an enormous number of bicycles or frequent redistribution, further improving the system service level is against bike-sharing firms' economic interest.

Therefore, we try to establish the link between the service level and the corresponding renewal rate for existing subscribers and the link between the service level and the opt-in rate for non-subscribers. We aim to identify the effect of the system performance and subscription fee on customer subscriptions, specifically the probability of riders purchasing the monthly subscription plan. In this section, We describe the overall methodology of our analysis, the construction of our variables, and the corresponding estimation results. We follow the order in which we construct rider characteristics, starting with the analysis of subscription renewals, and following with the new sign-ups.

3.4.1 Subscription Renewals

As mentioned in Section 3.3.3, existing subscribers face the renewal decision every 30 days and he/she can choose to renew anytime before the subscription expires. To capture the subscriber's perception of the service level in her decision-making process, we use equation 3.1

to calculate her sl on the day a renewal is made, if there is one, or on the expiration day if no renewal happens. For instance, consider a subscriber whose membership window is from day $t+1$ to day $t+30$, then there is a renewal event on day $t+30$. If the subscriber renewed her subscription on day $t+25$, then her sl will be calculated using data within day $t-4$ to $t+25$. If no renewal happened, then data within day $t+1$ to $t+30$ will be used. Note that for riders with none or few APP-related activities during the 30-day window (which means little system experience during the time), we use the market-average level sl to represent her perception of the service level.

While in their subscription period, existing subscribers can choose to turn on the auto-renewal so that their membership automatically gets extended for another month before it expires. Since subscribers can turn on the auto-renewal or turn it back off anytime they want, the auto-renew status signifies the willingness of the subscriber to extend her membership. It is also a decision (turn the auto-renewal on/off) that can be affected by subscribers' perception of service level.

Another important factor that affects subscription decisions is price. Meituan Bike provides a constant nominal price (25 CNY) for the monthly plan in Beijing but offers various types of discounts to different riders. During the timespan of our dataset, due to a citywide price experiment, riders in Beijing are randomly exposed to different discounts offered by the company and therefore selection bias is not a concern. The actual subscription fee is recorded for the converted events, in which existing subscribers choose to renew. However, for the non-converted events, the offered price can not be directly observed and needs the help of APP-related activities. Riders are exposed to their offered price only if they view the APP page of the subscription. For this reason, we consider the latest exposed price within the subscription window as the offered price for the corresponding renewal event. If there is no price exposure during the period, we use the market average exposed price.

After including some control variables and the time fixed effects, we consider the following logistic regression model:

$$\ln\left(\frac{Pr(\text{Renew}_{it}=1)}{1-Pr(\text{Renew}_{it}=1)}\right) = \alpha_0 + \alpha_1 sl_{it} + \alpha_2 \text{ Price}_{it} + \alpha_3 \text{ Is\_Auto\_Renew}_{it} + \lambda_{m(t)} + \gamma_{w(t)} + \beta \boldsymbol{X}_{it} + \epsilon_{it}, \tag{3.2}$$

where i and t denote the subscriber and date; $m(t) : t \to \{\text{Jan}, \text{Feb}, \ldots, \text{Aug}\}$ maps each date t to its corresponding month; $w(t) : t \to \{\text{Mon}, \text{Tue}, \ldots, \text{Sun}\}$ maps each date t to its corresponding day of the week; Renew_{it} indicates whether subscriber i extends her subscription for another month on day t; sl_{it} represents the service level rider i percepts on day t as previously defined; Price_i is the actual subscription price for subscriber i on day t; Is_Auto_Renew_i indicates whether user i is an auto-renew subscriber on day t. We control for the characteristics of the subscriber for each event, denoted by $\boldsymbol{X}_{it}$, which includes all the control variables: gender, age, the average number of APP-related activities per day in the previous week, and some other characteristics of users. Lastly, $\lambda_{m(t)}$ represents the month-of-the-year fixed effect and $\gamma_{w(t)}$ represents the day-of-the-week fixed effect. The estimation results are shown in Table 3.2.

As can be seen, subscription renewal probability decreases when the subscription fee increases. The odds of subscription renewal will be reduced by $1 - e^{0.1*(-2.43)} \approx 22\%$ if the subscription fee is 0.1 CNY (per day on average) higher. This makes perfect sense since a lower price means the user can gain higher utility from the subscription. More importantly, we see a significant positive impact of service level on renewals. The odds of subscription renewals can

Table 3.2: Estimation Results of Renewal Events

	Estimates	Std. Error
Service Level & Price		
Service Level	1.72^{***}	(0.0966)
Is_Auto_Renew	4.98^{***}	(0.0683)
Price	-2.43^{***}	(0.1098)
Rider Characteristics		
Female	-0.13^{***}	(0.0258)
Age: 30	0.23^{***}	(0.0373)
Age: 40	0.29^{***}	(0.0394)
Age: 50	0.41^{***}	(0.0482)
Age: 60	0.31^{***}	(0.0705)
Age: 70	0.45^{**}	(0.1619)
Time Fixed Effects		
Month: February	-0.15^{*}	(0.0703)
Month: March	0.29^{***}	(0.0637)
Month: April	-0.0175	(0.0607)
Month: May	0.04	(0.0598)
Month: June	-0.13^{*}	(0.0595)
Month: July	-0.0507	(0.0583)
Month: August	-0.17^{**}	(0.0573)
Day-of-Week: Monday	0.29^{***}	(0.0497)
Day-of-Week: Tuesday	0.43^{***}	(0.0497)
Day-of-Week: Wednesday	0.46^{***}	(0.0464)
Day-of-Week: Thursday	0.4^{***}	(0.0476)
Day-of-Week: Saturday	-0.2^{***}	(0.0492)
Day-of-Week: Sunday	-0.32^{***}	(0.0482)
AIC	37540	
Observations	43,970	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

be boosted by $e^{(0.1*1.72)} \approx 119\%$ if the user's perceived service level increases by 0.1. Another factor that indicates higher odds is Is_Auto_Renew. Subscribers who turned on the auto-renewal have $e^{4.98} \approx 145\%$ higher odds of extending their memberships.

Table 3.2 also shows that female subscribers and younger subscribers (age 20 and below) are less likely to renew, given other conditions the same. Subscribers are also less willing to renew in winter and summer when the temperature is usually not biking-friendly. In addition, compared to the weekend, subscribers are more willing to renew during weekdays, when most of the need for rides is generated.

3.4.1.1 Auto-renewal

From table 3.2, we observe that the status of auto-renewal has a huge impact on the renewal decision. Recall that turning the auto-renewal on/off is a decision that needs to be made by subscribers during the membership period. Therefore, we consider the following logistic regression model to capture the impact of service level on the auto-renewal decision:

$$\ln\left(\frac{Pr(\text{Is\_Auto\_Renew}_{it}=1)}{1-Pr(\text{Is\_Auto\_Renew}_{it}=1)}\right)=\alpha_0+\alpha_1 sl_{it}+\lambda_{m(t)}+\gamma_{w(t)}+\beta\boldsymbol{X}_{it}+\epsilon_{it}. \tag{3.3}$$

The estimation results are shown in Table 3.3.

As presented in Table 3.3, the odds of turning on auto-renewal can be boosted by $e^{(0.1*1.29)} \approx 114\%$ if the user's perceived service level increases by 0.1. Also, female subscribers and younger subscribers are less likely to turn on auto-renewal, given other conditions the same.

Consequently, service level impacts the renewal probability through two channels: directly affecting the renewal probability and indirectly by affecting the probability of becoming an auto-renew subscriber.

3.4.2 New Sign-ups

Now we consider the sign-up events for the non-subscribers. When a non-subscriber purchases a monthly pass, a sign-up event is observed. The variables can be constructed the same way as in the renewal events. To construct the sign-up event which ends up with no purchase, we

Table 3.3: Estimation Results of Auto-renewal

	Estimates	Std. Error
Service Level	1.29***	(0.0844)
Rider Characteristics		
Female	-0.15***	(0.0023)
Age: 30	0.18***	(0.0335)
Age: 40	0.18***	(0.0356)
Age: 50	0.19***	(0.0439)
Age: 60	0.45***	(0.0596)
Age: 70	0.83**	(0.12)
Time Fixed Effects		
Month: February	1.51***	(0.1151)
Month: March	1.74***	(0.1109)
Month: April	1.88***	(0.1079)
Month: May	2.18***	(0.1066)
Month: June	2.3***	(0.1061)
Month: July	2.16***	(0.106)
Month: August	1.98***	(0.1056)
Day-of-Week: Monday	0.07	(0.0426)
Day-of-Week: Tuesday	0.06	(0.0432)
Day-of-Week: Wednesday	-0.11**	(0.0417)
Day-of-Week: Thursday	-0.06	(0.0422)
Day-of-Week: Saturday	-0.1*	(0.0421)
Day-of-Week: Sunday	0.02	(0.0419)
AIC	46996	
Observations	43,970	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

consider that all the non-subscribers who are active face a purchase/no purchase decision for the monthly pass. An active non-subscriber is defined as a rider who is currently not in membership and has at least seven APP-related activities in the past seven days (on average at least one time per day). Using this rule and following similar variable construction procedures discussed in Section 3.4.1, we consider the following logistic regression model:

$$\ln\left(\frac{Pr(\text{Sign-up}_{it} = 1)}{1 - Pr(\text{Sign-up}_{it} = 1)}\right) = \alpha_0 + \alpha_1 sl_{it} + \alpha_2 \,\text{Price}_{it} + \lambda_{m(t)} + \gamma_{w(t)} + \beta \boldsymbol{X}_{it} + \epsilon_{it}, \quad (3.4)$$

Table 3.4: Estimation Results of New Sign-up Events

	Estimates	Std. Error
Service Level & Price		
Service Level	1.96***	(0.0516)
Price	-1.28***	(0.0811)
Rider Characteristics		
Female	0.31***	(0.0164)
Age: 30	0.12***	(0.0222)
Age: 40	0.07**	(0.0242)
Age: 50	-0.01	(0.0318)
Age: 60	-0.11*	(0.0476)
Age: 70	-0.1	(0.1132)
Time Fixed Effects		
Month: February	0.1*	(0.0484)
Month: March	0.13**	(0.0449)
Month: April	0.003	(0.0435)
Month: May	-0.01	(0.043)
Month: June	-0.07	(0.0435)
Month: July	0.01	(0.0426)
Month: August	0.02	(0.0431)
Day-of-Week: Monday	0.18***	(0.029)
Day-of-Week: Tuesday	0.12***	(0.0297)
Day-of-Week: Wednesday	0.03	(0.0308)
Day-of-Week: Thursday	0.002	(0.0312)
Day-of-Week: Saturday	-0.19***	(0.0326)
Day-of-Week: Sunday	-0.29***	(0.0329)
AIC	120115	
Observations	317,782	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

where Sign-up$_{it}$ indicates whether rider i purchases the monthly pass on day t; the other variables are as previously defined. The estimation results are shown in Table 3.4.

Similar to the renewal events, the odds of subscription are decreasing in the subscription fee, and we see a significant positive impact of service level on the new sign-up probability. The odds of new sign-up can be boosted by $e^{(0.1*1.96)} \approx 122\%$ if the user's perceived service level increases by 0.1. In addition, compared to the weekend, riders are more willing to sign up during weekdays, when most of the need for rides is generated.

In contrast to the renewal events, Table 3.4 shows that female riders and younger riders are more likely to purchase the monthly membership, given other conditions the same.

3.5 Counterfactual Analysis

In this section, we first describe the framework we use for our counterfactual analysis in Section 3.5.1. Then, based on our estimations in Section 3.4, we conduct two counterfactual experiments in Section 3.5.2 and 3.5.3 to study how bicycle fleet size and price will shape the evolution of subscriptions and affect the company profit. The first counterfactual analysis examines the impact of the bike fleet size on service level and subscription probabilities in various events, which in the end translates into subscription retentions and company profit. We demonstrate that albeit a larger fleet size implies a higher service level and makes the membership plan more attractive, deploying more bikes incurs larger purchasing and maintenance costs, and has a decreasing marginal return on system performance. More importantly, our framework would allow bikeshare companies to assess the financial impact of fleet size adjustment on subscription conversions and company profit under various market conditions. In the second analysis, we examine several price points for the monthly pass and demonstrate that the monthly membership is underpriced. Our framework allows bikeshare companies to have a better understanding of how their pricing strategy going to affect the subscriptions and find the price that maximizes the company profit.

3.5.1 Setup

In order to understand how different factors affect company profit, we need to keep track of riders' subscription behavior. Recall that there are two streams of subscription revenue: non-subscriber sign-ups and existing subscriber renewals. Specially, we have:

$$\begin{aligned}\text{Total\_Revenue} &= \sum_{t=1,\dots,T} \text{Revenue\_Renewal}_t + \text{Revenue\_New\_Sign-up}_t \\ &= \sum_{t=1,\dots,T} \left(\sum_{i=1,\dots,RE_t} p_{it} \cdot \mathbb{1}\{\text{Renew}_{it} = 1\} + \sum_{i=1,\dots,NE_t} p_{it} \cdot \mathbb{1}\{\text{Sign-up}_{it} = 1\} \right). \end{aligned} \tag{3.5}$$

Here, p_{it} is the subscription fee for rider i on day t; RE_t represents the number of existing subscribers that face the renewal decision on day t; NE_t is the number of non-subscribers that face the sign-up decision on day t; the probability of renewal and sign-up are both functions of service level sl, price p_{it}, and other variables as specified in Section 3.4. Note that the service level is dependent on both the bike fleet size and the number of subscribers. A larger fleet size implies a higher service level. However, since a higher service level may also attract more riders to subscribe or stay in the membership, it leads to a heavier need of rides for the system, which mChapter 2ay then lower the service level. Next, to capture this interplay and dynamics between the service level and the number of subscribers, we establish the relationship between a given fleet size and number of subscribers to the resulting service level.

Bicycle Fleet Size to Service Level We assume that the system service level is a function of the ratio between fleet size and the number of APP-related events. This ratio, which we denote as "BA_ratio", can be considered as a proxy for number of bikes per demand. Higher BA_ratio

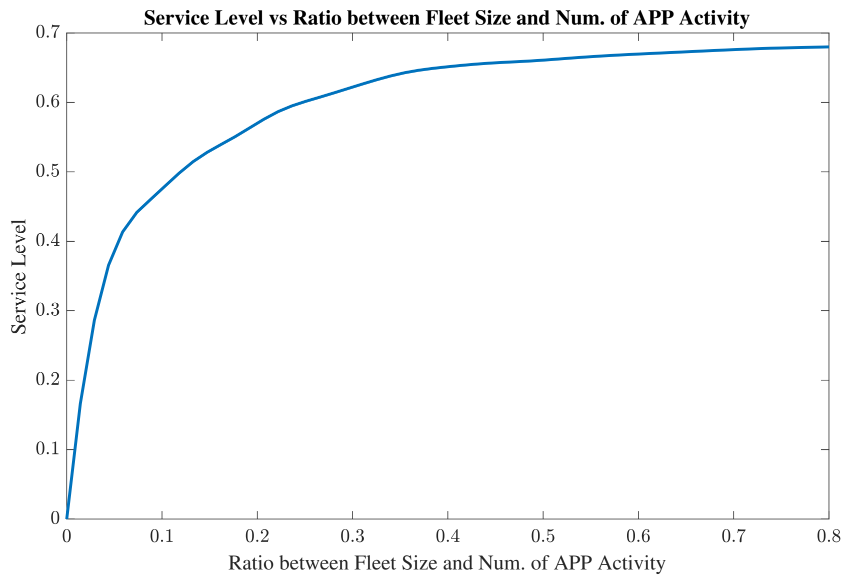

Figure 3.2: Service Level vs BA_ratio

implies that on average, each rider is served by more bicycles. Based on the counterfactual framework proposed in Chapter 2, we conduct the simulation using various BA_ratio and get the function $sl = f(\text{BA\_ratio})$, i.e., the functional relationship between system service level and BA_ratio. This relationship is shown in Figure 3.2.

As can be observed, while a higher bicycle per activity (BA_ratio) ensures a higher system service level, the marginal effect of enlarging fleet size is decreasing. This is consistent with our results in Chapter 2, that purely flooding the market is not an efficient way to improve system performance.

Subscription Evolution Now Let's look at how different factors affect the revenue by affecting subscription probabilities. Denote S_t as the number of riders who renewed on day t and NS_t as the number of new sign-ups on day t. Therefore, we have the following relationships:

$$TS_{t+30} = \sum_{t'=t,\dots,t+30} S_{t'} + NS_{t'}, \tag{3.6}$$

$$RE_{t+30} = S_t + NS_t, \tag{3.7}$$

$$S_t = \sum_{i=1,\dots,RE_t} \mathbb{1}\{\text{Renew}_{i,t} = 1\}, \tag{3.8}$$

$$NS_t = \sum_{i=1,\dots,NE_t} \mathbb{1}\{\text{Sign-up}_{i,t} = 1\}, \tag{3.9}$$

for $t \geq 30$. We assume that NE_t is independent of price or service level and is given as constructed from the data. In addition, according to Section 3.4,

$$Pr(\text{Renew}_{it} = 1) = f_r(sl_{it},\ \text{Price}_{it},\ \text{Is\_Auto\_Renew}_{it}(sl_{it}, \lambda_{m(t)}, \gamma_{w(t)}, \boldsymbol{X}_{it}),\ \lambda_{m(t)},\ \gamma_{w(t)},\ \boldsymbol{X}_{it}), \tag{3.10}$$

$$Pr(\text{Sign-up}_{it} = 1) = f_s(sl_{it},\ \text{Price}_{it},\ \lambda_{m(t)},\ \gamma_{w(t)},\ \boldsymbol{X}_{it}), \tag{3.11}$$

Therefore, given bike fleet size and subscription fee, S_t, NS_t, and RE_t can be updated sequentially. We summarize this evolution of subscriptions in Algorithm 3.

3.5.2 Bicycle Fleet Size

Deploying more bicycles in the market is one of the most straightforward and commonly used ways to improve system performance and rider satisfaction. However, it also incurs additional costs for the company. Besides the purchasing cost at the beginning, a larger fleet size implies much more maintenance cost in the daily system operations. Chapter 2 shows that de-

Algorithm 3:

input : Fleet size N, subscription fee p_{it}, RE_t $(t = 1, \dots, 30)$, NE_t $(t = 1, \dots, T)$, rider characteristic distribution.

for $t \in \{1, \dots, T\}$ **do**

  Calculate BA_ratio$_t$ = $\frac{N}{app\_s \cdot \sum_{t'=t,\dots,t+29} RE_{t'} + app\_ns \cdot (M - \sum_{t'=t,\dots,t+29} RE_{t'})}$,

    $sl_t = f(\text{BA\_ratio}_t)$;

  Update $RE_t = S_{t-30} + NS_{t-30}$;

  for $i \in \{1, \dots, RE_t\}$ **do**

    Simulate rider i's characteristics;

    Simulate Is_Auto_Renew$_{it}$;

    Simulate Renew$_{it}$;

    if Renew$_{it} = 1$,

    then Update $S_{t+30} = S_{t+30} + 1$, Revenue $=$ Revenue $+ p_{it}$;

  end for

  for $i \in \{1, \dots, NE_t\}$ **do**

    Simulate rider i's characteristics;

    Simulate Sign-up$_{it}$;

    if Sign-up$_{it} = 1$,

    then Update $NS_{t+30} = NS_{t+30} + 1$, Revenue $=$ Revenue $+ p_{it}$;

  end for

end for

ploying more bikes is not an efficient approach to improve system efficiency. In this section, we vary bicycle fleet size N, try to understand the dynamics between fleet size and subscription, and how does fleet size affect the system service level and company profit. We present the impact of fleet size on company profit in Figure 3.3.

As shown in Figure 3.3, company profit increases rapidly with larger fleet sizes when the fleet size is small. This is because the additional number of bicycles has a large impact on service level. However, the marginal improvement on service level is decreasing, as can be observed in Figure 3.3. As a result, after the fleet size reaches a certain level, the marginal benefit of more bicycles can not offset the newly incurred costs. Hence the profit goes down if further increasing the fleet size.

Figure 3.3 also shows that the optimal fleet size is around 120,000 bicycles. Therefore,

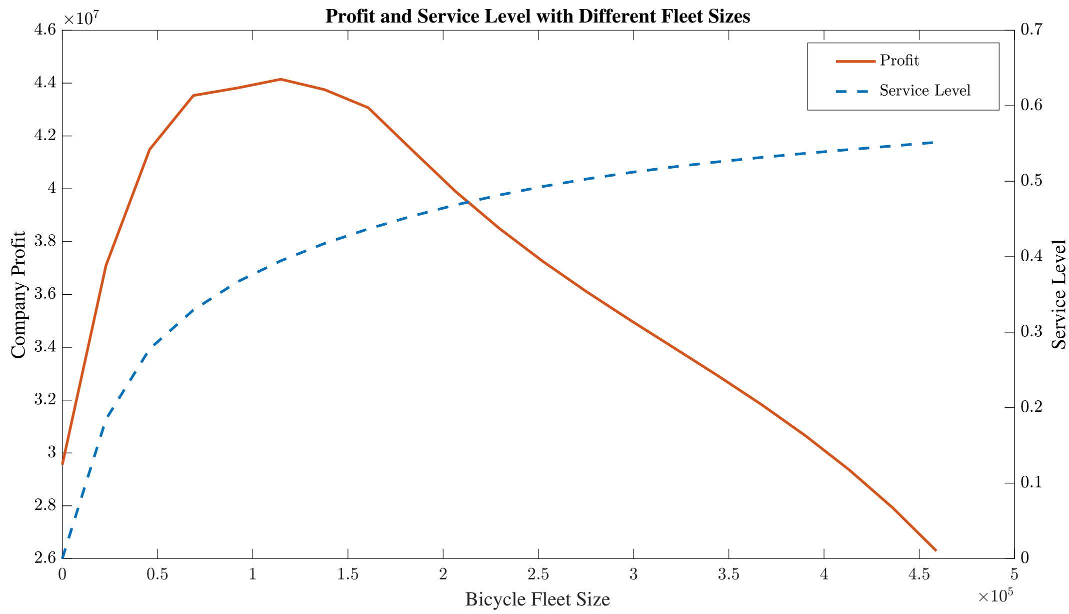

Figure 3.3: The impact of fleet size on company revenue and service level.

given other conditions the same, downsizing the fleet saves more money than the revenue loss from the subscription in this case. Our study provides important practical guidance for bikeshare companies to evaluate and adjust the bicycle fleet size in order to maximize profit. The framework can also be easily applied to other dockless bike sharing systems in other cities to evaluate operational policies.

3.5.3 Price

Bikesharing firms tend to underprice in the early stage of the business, and often offer promotions and discounts, trying to attract more riders. As the goal switches from drawing riders to achieving sustainable profitability, the choice of price is extremely important. Companies no longer want to burn money and provide free rides to customers, while they are also afraid of losing riders if pricing too aggressive. To better understand how subscription price will shape subscription behavior and revenue, in this section we examine some fixed prices for the monthly

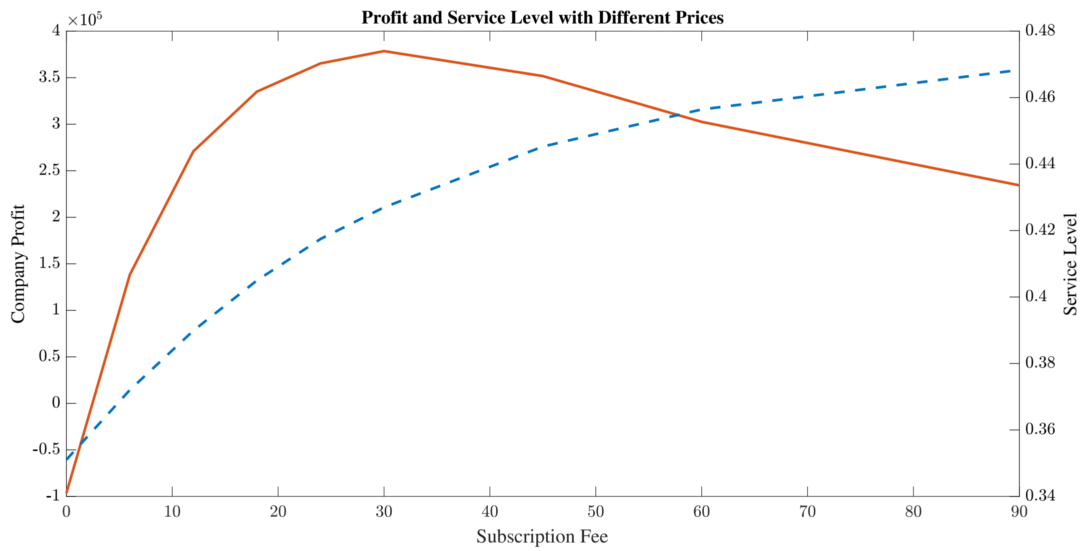

Figure 3.4: The impact of fleet size on company revenue and service level.

pass and check the impact it has on service level and profit. The results are displayed in Figure 3.4.

Consistent with the results in Section 3.4, when offering higher prices, fewer riders will convert and become subscribers. This change implies a better service level since the system will have less demand. Recall that the nominal price is 25 CNY and the average effective price during the timespan is about 15 CNY (0.49*30). Figure 3.4 says that it is safe for the company to increase the price, and they can achieve the maximized profit if raising the price to around 30 CNY. That is about 25% profit increase compared with the current pricing strategy.

3.6 Concluding Remarks

Using data from a dockless bike sharing system in Beijing, we study the subscription behavior and its relationship with the service level and price. Bikeshare companies have been through intense competition in the early stage of the business, where they kept flooding the market with an excessive number of bicycles, aiming to improve system service level and rider satisfaction; offering low prices and discounts, trying to attract more riders and retain them. Now the cash-burning stage is over, surviving players have different targets and are thinking about sustainability and profitability.

In this work, we develop an econometric model to study the subscription behavior for both existing subscribers and new sign-ups. We build up the functional relationship between the service level and system demand level and reveal the dynamics and interplays among subscription, system demand, and service level. This helps us to recover how the number of subscribers evolves over time. Based on all the estimation results and functional relationships, we then construct the empirical framework and straighten out the relationship between company profit and bicycle fleet size/subscription price.

Our results confirm that the marginal benefit of deploying a larger bicycle fleet is decreasing. After a certain point, the cost of purchasing and maintaining a large fleet outplays the increased revenue that can be generated from higher subscription numbers. Therefore, the company should be more cautious in determining and adjusting the bicycle fleet size, and resort to other approaches (e.g., bike redistribution under a reasonable budget) that can increase the service level without incurring too many costs.

Another lever that bikeshare firms used to underplay is pricing. Lower prices and discounts attract more riders and are often used when the business is growing. While it is not a profitable or sustainable way to operate a business, firms are reluctant to price more aggressively, worried about losing riders. Our examination demonstrates that the current price is too low and raising the price properly can achieve about a 25% profit increase. Our results also show the value of understanding riders' sensitivity to price and how to use it to better their operational decisions and accomplish better financial results. The counterfactual framework we proposed can be utilized in various policy evaluations and provides important insights and recommendations to the bikeshare companies and regulators.

One caveat of this study is that the timespan (January to August 2021) of the data we use is after the outbreak of the COVID-19 pandemic. Although during the time Beijing was only mildly impacted by the pandemic (as a major city with over 20 million population, Beijing only had around 120 cases in total during the eight-month period according to JHU CSSE COVID-19 Data), it may still have had an impact on commuters' mindset and mobility behavior. A possible future research direction can be to analyze pre-pandemic subscription behavior and make a comparison. Another potential direction is to look at and optimize the bike redistribution strategies by involving the relationship between service level and rider subscription behavior.

References

[1] Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, 12(4):3, 2009.

[2] NACTO. Shared micromobility in the U.S.: 2019. Available at `https://nacto.org/shared-micromobility-2019/`, 2020.

[3] PYMNTS. SoBi and the smarter way to do bike sharing. Available at `https://www.pymnts.com/news/retail/2017/sobi-and-the-smarter-way-to-do-bike-sharing/`, 2017.

[4] The State Information Center. Annual report on the development of shared mobility in China, 2019.

[5] NACTO. Shared micromobility in the U.S.: 2017. Available at `https://nacto.org/bike-share-statistics-2017/`, 2018.

[6] Divya Singhvi, Somya Singhvi, Peter I Frazier, Shane G Henderson, Eoin O'Mahony, David B Shmoys, and Dawn B Woodard. Predicting bike usage for New York City's bike sharing system. In *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, 2015.

[7] Wafic El-Assi, Mohamed Salah Mahmoud, and Khandker Nurul Habib. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, 44(3):589–613, 2017.

[8] Chong Yang Goh, Chiwei Yan, and Patrick Jaillet. Estimating primary demand in bike-sharing systems. *Available at SSRN 3311371*, 2019.

[9] Konstantina Mellou and Patrick Jaillet. Dynamic resource redistribution and demand estimation: An application to bike sharing systems. *Available at SSRN 3336416*, 2019.

[10] Ashish Kabra, Elena Belavina, and Karan Girotra. Bike-share systems: Accessibility and availability. *Management Science*, 66(9):3803–3824, 2020.

[11] Pu He, Fanyin Zheng, Elena Belavina, and Karan Girotra. Customer preference and station network in the London bike-share system. *Management Science*, 67(3):1392–1412, 2021.

[12] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 63(4):841–890, 1995.

[13] Guangyu Cao, Ginger Zhe Jin, and Li-An Zhou. Market expanding or market stealing? Platform competition in bike-sharing. *National Bureau of Economic Research*, Vol. 24938, 2018.

[14] Junhong Chu, Yige Duan, Xianling Yang, and Li Wang. The last mile matters: Impact of dockless bike sharing on subway housing price premium. *Management Science*, 67(1):297–316, 2021.

[15] Ruomeng Cui, Meng Li, and Jingbo Zhou. Last-mile commute: Impact of bike-sharing on restaurants. *Available at SSRN 3596533*, 2020.

[16] Daniel Chemla, Frédéric Meunier, and Roberto Wolfler Calvo. Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, 10(2):120–146, 2013.

[17] Mauro Dell'Amico, Eleni Hadjicostantinou, Manuel Iori, and Stefano Novellani. The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, 45:7–19, 2014.

[18] Eoin O'Mahony, Shane G Henderson, and David B Shmoys. Bike sharing. Technical report, Cornell working paper, 2016.

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

[20] Supriyo Ghosh, Pradeep Varakantham, Yossiri Adulyasak, and Patrick Jaillet. Dynamic repositioning to reduce lost demand in bike sharing systems. *Journal of Artificial Intelligence Research*, 58:387–430, 2017.

[21] Yingxiang Yang, Carlos Herrera, Nathan Eagle, and Marta C González. Limits of predictability in commuting flows in the absence of data for calibration. *Scientific Reports*, 4:5662, 2014.

[22] Chris Veness. Calculate distance, bearing and more between latitude/longitude points. *Movable Type Scripts*, pages 2002–2014, 2010.

[23] Josh Ye. Why China's bike-sharing boom is causing headaches. Available at `https://www.scmp.com/news/china/society/article/2085751/chinas-bike-sharing-boom-puts-pressure-planners`, 2017.

[24] Eliot Brown. Dockless bike share floods into U.S. cities, with rides and clutter. Available at `https://www.wsj.com/articles/dockless-bike-share-floods-into-u-s-cities-with-rides-and-clutter-1522076401`, 2018.

[25] Alan Taylor. The bike-share oversupply in China: Huge piles of abandoned and broken bicycles. *The Atlantic*, 22, 2018.

[26] Hui Jiang. Chinese cities aim to rein in bike-sharing boom. Available at `https://www.wri.org/insights/chinese-cities-aim-rein-bike-sharing-boom`, 2018.

[27] Spencer Yang. A closer look into bike-sharing in China and its future. Available at `https://www.techinasia.com/talk/bike-sharing-china-future`, 2017.

[28] Luz Lazo. Hey, you can't park there! Dockless bike-share bikes ending up in inappropriate places. Available at `https://www.washingtonpost.com/news/dr-gridlock/wp/2017/10/05/abandoned-vandalized-and-illegally-parked-bike-share-bikes-now-a-d-c-problem/`, 2017.

[29] Susan A Shaheen, Adam P Cohen, and Elliot W Martin. Public bikesharing in North America: early operator understanding and emerging trends. *Transportation research record*, 2387(1):83–92, 2013.

[30] Susan A Shaheen, Stacey Guzman, and Hua Zhang. Bikesharing in europe, the americas, and asia: past, present, and future. *Transportation Research Record*, 2143(1):159–167, 2010.

[31] Luigi Dell'Olio, Angel Ibeas, and Jose Luis Moura. Implementing bike-sharing systems. *Proceedings of the Institution of Civil Engineers*, 164(2):89, 2011.

[32] David William Daddio. The impact of weather conditions on bikeshare trips in washington, dc. 2012.

[33] Kyle Gebhart and Robert B. Noland. Maximizing bicycle sharing: An empirical analysis of capital bikeshare usage. *Proceedings of the Institution of Civil Engineers*, 41(6):1205–1225, 2014.

[34] Elliot Fishman, Simon Washington, and Narelle Haworth. Bike share: a synthesis of the literature. *Transport reviews*, 33(2):148–165, 2013.

[35] Elliot Fishman. Bikeshare: A review of recent literature. *Transport Reviews*, 36(1):92–113, 2016.

[36] Adithya Pattabhiramaiah, Eric Overby, and Lizhen Xu. Spillovers from online engagement: how a newspaper subscriber's activation of digital paywall access affects her retention and subscription revenue. *Management Science*, 68(5):3528–3548, 2022.

[37] Peter J Danaher. Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Science*, 21(2):119–138, 2002.

[38] Longxiu Tian and Fred M Feinberg. Optimizing price menus for duration discounts: A subscription selectivity field experiment. *Marketing Science*, 39(6):1181–1198, 2020.

[39] John Birge, Ozan Candogan, Hongfan Chen, and Daniela Saban. Optimal commissions and subscriptions in networked markets. *Manufacturing & Service Operations Management*, 23(3):569–588, 2021.

[40] Ian A Kash, Peter B Key, and Spyros I Zoumpoulis. Optimal pricing and introduction timing of technology upgrades in subscription-based services. *Operations Research*, 2022.

[41] Yunke Mai and Bin Hu. Optimizing free-to-play multiplayer games with premium subscription. *Management Science*, 2022.