ESSAYS ON FOR-HIRE VEHICLE MARKETS

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Kin-ping Jeremy Wong, M.A.

Washington, DC
April 26, 2019
Copyright © 2019 by Kin-ping Jeremy Wong
All Rights Reserved
ESSAYS ON FOR-HIRE VEHICLE MARKETS

Kin-ping Jeremy Wong, M.A.

Thesis Advisor: John P. Rust, Ph.D.

ABSTRACT

This dissertation consists of two essays of empirical studies on the for-hire vehicle market. In the first chapter, I present a dynamic spatial search and match model to study the effects of matching improvements including e-hail dispatch platforms, and flexible fare and leasing schedules on drivers' income, utilization rates, and consumer surplus for passengers. With the aid of real time price and waiting time data collected from Uber, I estimate the price and waiting time elasticities on the demand side to predict the response on net demand for taxicabs under different regimes. Counterfactual results indicate that using a flexible leasing schedule increases the number of completed trips by 3,710 and improved average shift earnings by $7.2. Waiving the current e-hail booking fee and implementing a universal e-hail dispatcher for the city result in higher usage of e-hail, generating an additional consumer surplus of $0.15 per commuter per day, aggregating to $25.32 million per year.

In the second chapter, I study the e-hail taxicab market operated by competing digitized ride hailing dispatchers. Dynamic pricing in ride-sharing platforms has been shown to filter excess demand and incentivize supply in busy times and locations to allowing more efficient allocation of vehicles to serve riders in need. I found evidence of spatial reallocation of Uber drivers under price surges in real time data collected from Uber’s application programming interface, suggesting similar pricing schemes can bring about gains for e-hail taxi market even if “gig-economy” shift schedules are not available for taxicab drivers. I extend the the structural model of taxicab market presented in chapter 1 to study the welfare effects under different pricing strategies.
and industrial organization in the e-hail taxi market. Simulation experiments suggest that e-hail markets exhibit positive network externalities. Social surplus is highest when the e-hail taxi market is operated by a monopolist. Under all settings considered, surge pricing generates a mutual gain for riders and platform. Spatial pricing provides gains for platforms at the expense of passengers except when operated by a social welfare maximizing monopolist.

INDEX WORDS: Taxicab, Spatial Competition, Network Externalities
DEDICATION

Dedicated to my parents, my sister Anne and my girlfriend Karen.
ACKNOWLEDGMENTS

I am indebted to the invaluable advice and encouragements from my thesis advisor, John Rust, who also generously shares his fast computing machine that greatly helps my progress. I also thank my dissertation committee members Axel Anderson and James Albrecht for their constructive comments.

I thank Renee Brown, Nick Buchholz, Ginger Jin, Alessandro Lizzeri, Bertel Schjerning, Andrew Sweeting, Francis Vella and participants of Georgetown seminars for their helpful comments and discussions.
# Table of Contents

Acknowledgments ................................................................. vi

CHAPTER

1 E-hail, Flexible Fare and Leasing: Improving the New York Taxicab Market 1
  1.1 Introduction ............................................................... 1
  1.2 Data Description ......................................................... 12
  1.3 The Model ................................................................. 25
  1.4 Estimation ................................................................. 41
  1.5 Results ................................................................... 46
  1.6 Counterfactual Experiments .......................................... 56
  1.7 Conclusion ............................................................... 66

2 Dynamic Pricing and Competition in the New York City For-Hire Vehicle Industry ................................................................. 68
  2.1 Introduction ............................................................... 68
  2.2 Empirical Findings in Data ........................................... 74
  2.3 The Model ................................................................. 79
  2.4 Competitive Equilibrium .............................................. 86
  2.5 Computation and Calibration ....................................... 94
  2.6 Counterfactual Simulations ......................................... 103
  2.7 Conclusion ............................................................... 120

APPENDIX

A Simulated Generalized Method of Moments Estimates .................. 122

B An Experiment on Carpooling ............................................ 124

C A Discussion on the Parametric Forms of Matching Functions .... 126
  C.1 Street-Hail Matching Function .................................... 126
  C.2 E-hail Matching Function .......................................... 129

D A Discussion on Equilibrium Convergence ................................. 132
  D.1 Continuity of $\Gamma$ .................................................. 132
  D.2 Successive Belief Iteration in a Static Matching Game ........ 133
  D.3 Dynamic Matching Game ............................................ 138

E Details on Demand Data Imputation .................................... 140
  E.1 Estimation ............................................................... 140
# List of Tables

1.1 Descriptive statistics for yellow cabs trips. ................................ 13
1.2 Descriptive Statistics for green cab trips. ................................. 14
1.3 Number of pickups for each for-hire vehicle in June 2016. ............ 16
1.4 Outcome of e-hail requests. .................................................. 20
1.5 Location of requests completed by yellow cabs by borough, 2013-2014. 21
1.6 Descriptive statistics for API data. .......................................... 23
1.7 Estimated time a cab stays in zone where it dropped off a passenger ($\hat{x}$). 43
1.8 Shift descriptive statistics on model fit (2013). ......................... 47
1.9 Estimates of demand parameters. .......................................... 54
1.10 UberTAXI waiting time data. ................................................. 55
1.11 Equilibrium outcomes across percentage of street hail demand substitut- ed. .......................................................... 58
1.12 Earnings for e-hail vs non e-hail drivers. ................................. 58
2.1 Regression of waiting times on average neighbor surge. ................ 77
2.2 Demand estimates. ............................................................. 98
2.3 Estimates of waiting time function. ........................................ 101
2.4 Equilibrium for benchmark cases. ........................................... 104
2.5 Optimal monopolist pricing. ................................................ 106
2.6 Monopolist equilibrium with optimal dynamic pricing. ................. 106
2.7 Pure strategy payoffs. ......................................................... 112
2.8 Outcomes of duopoly equilibrium under surge pricing competition. . 112
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9          Optimal spatial pricing multipliers.</td>
<td>116</td>
</tr>
<tr>
<td>2.10         Outcomes of duopoly equilibrium under spatial pricing competition.</td>
<td>116</td>
</tr>
<tr>
<td>A.1          Simulated generalized method of moments estimates.</td>
<td>123</td>
</tr>
<tr>
<td>B.1          Percentage increase in matches for each given % of street hail demand substituted.</td>
<td>125</td>
</tr>
<tr>
<td>E.1          Statistics of trip characteristics in imputed data.</td>
<td>143</td>
</tr>
<tr>
<td>F.1          Waiting times outside Manhattan.</td>
<td>156</td>
</tr>
<tr>
<td>G.1          Equilibrium outcome.</td>
<td>158</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 Yellow zone and heat map of passenger pick ups. ........................................ 13  
1.2 Distribution of shift start time. ............................................................... 15  
1.3 Average daily number of pickups for yellow (bold) and green cabs (dotted). ............................................................... 15  
1.4 Taxi and app for-hire vehicle pickups. ..................................................... 17  
1.5 Market shares of taxicab and other for-hire vehicle. ............................ 18  
1.6 Total (solid) and completed (dotted) e-hail requests in the year 2016. 21  
1.7 Total e-hail requests from 2013 June to 2014 December. ...................... 22  
1.8 Average surge multiplier in Manhattan (yellow) and other boroughs (green). ........................................................................................................ 24  
1.9 Average surge multiplier and waiting time .............................................. 24  
1.10 Convergence of beliefs of street hail ride probabilities. ........................... 39  
1.11 Convergence of beliefs of e-hail ride probabilities. ................................. 39  
1.12 Simulated and true distributions of shift earnings. ................................. 48  
1.13 Simulated and true distributions of utilization rates. ............................... 48  
1.14 Simulated and true street hail matches. ................................................. 49  
1.15 Simulated and true e-hail matches. ......................................................... 50  
1.16 Value function across all locations. ......................................................... 52  
1.17 Aggregated values of two 12-hour shifts. .............................................. 53  
1.18 Distribution of UberTAXI waiting time. ................................................... 55  
1.19 Street hail (solid) and UberTAXI (dotted) waiting time in Manhattan. 56
1.20 Simulated waiting times for street hail taxicabs (solid) vs. Uber (dotted) in Manhattan. ................................................. 60
1.21 Simulated waiting times for taxicabs (solid) vs. Uber (dotted) in Manhattan under universal dispatcher. ................................. 61
1.22 5th and 95th percentiles of simulated total expected surplus in a 12-hour shift (by shift starting hour). ........................................ 64
1.23 Active cabs (solid: flex leasing, dotted: current). ........................ 65
1.24 Completed trips (solid: flex leasing, dotted: current). .................... 66
2.1 Heat maps of average surge multiplier (left) and waiting time (right). 74
2.2 Average surge multiplier against average waiting time across all locations. 75
2.3 Clusters of neighboring zones in Manhattan. .............................. 77
2.4 Coefficients of average neighbor surge for all locations in Manhattan. 78
2.5 Logistics of the model. ...................................................... 80
2.6 Structure of competitive equilibrium. ........................................ 89
2.7 Three-location setup in the model. ........................................... 95
2.8 Passenger and cab arrivals at each location. ............................... 96
2.9 Average Uber surge multiplier and waiting time. .......................... 97
2.10 Simulated and fitted waiting times. ........................................ 100
2.11 Belief convergence for benchmark case. .................................... 102
2.12 Supply and demand for benchmark monopolist (left) and duopoly (right) cases. ................................................................. 105
2.13 Supply and demand under monopolist optimal dynamic pricing solution.110
2.14 Waiting times and request fulfillment rates under monopolist optimal dynamic pricing solutions. ........................................... 111
2.15 Supply and demand under surge pricing duopoly competition. ...... 114
Chapter 1

E-hail, Flexible Fare and Leasing: Improving the New York Taxicab Market

1.1 Introduction

The taxicab market in New York City is an example of a matching market with search frictions: there is an undersupply in areas of lower expected revenues. Search frictions arising from spatial mismatch of supply and demand is acknowledged by government agencies and the literature. Seminal works by Lagos (2000, 2003) and Buchholz (2018) provide theoretical and empirical evidence of search frictions: decentralized spatial search decisions of cab drivers result in equilibria that simultaneously exhibit vacant cabs and unserved passengers.

Recent entrants in the market, peer-to-peer ridesharing services (e.g. Uber) offered a digitized matching technology and are subject to less pricing and shift schedule restrictions, allowing them to gain a price and waiting time advantage over taxis. In this paper, I study the effects of three changes designed to reduce search frictions and improve the taxicab market: using a centralized dispatch platform to match passengers and cabs (e-hail), employ a flexible taxi fare rate and a flexible leasing schedule. Effects of policy regimes are evaluated in terms of drivers’ income, utilization rates of drivers and gains in consumer surplus.

The for-hire vehicle market in New York City is regulated by the Taxicab and Limousine Commission (thereafter TLC). The for-hire vehicle market consists of
taxicabs (yellow and green), traditional for-hire vehicles such as black cars, luxury
limousines and ridesharing services. Of these, taxicabs face the most restrictive entry
and price regulations. To legally operate in New York City, minifleets and individuals
are required to own a medallion licensed by TLC. The medallion system was intro-
duced as a mechanism to guarantee the quality of ridehailing service providers and
control the number of cabs. As of 2018, the number of medallions is capped at 13,587.
Operational restrictions such as fixed fare rates and minimum number of nine-hour
shifts per medallion are also imposed on taxicabs.

Around 2010, new technologies have emerged to reduce search frictions in the
for-hire vehicle industry. Most of these attempts are technologies to help digitize the
matching of passengers and cabs with smart phone applications. Most of the enhance-
ment comes from newcomers in the past decade: peer-to-peer ride-sharing platforms
such as Uber, Lyft and Via. By June 2016, Uber, Lyft and Via accounted for 48.77%
in the total market share in the for-hire vehicle market New York City. Taxicabs also
attempted using centralized dispatchers on smart phone applications (e-hail there-
after) to match passengers and vacant cabs. E-hail taxi apps have not contributed to
a significant share of all taxicab rides. As of year 2015, there were fewer than 7,900
e-hail requests in an average day (around 2% of daily total completed trips), and the
trip completion rate is lower than 60%. E-hail request and response data provided by
TLC and waiting time data collected from one of the e-hail taxi provider (UberTAXI)
suggest low completion rates were likely caused by prolonged waiting times and a low
response rate in locations at which e-hail requests are densest. While this may be a
result of endogenous spatial search decisions by drivers, a low number of drivers per
dispatcher (network strength) is also a likely cause. Another reason for the low e-hail
request rate is that under presence of other ride sharing alternatives, the regular taxi
fare plus the $2 booking fee exceeds passengers’ willingness-to-pay to pay for e-hail
cabs. In one of the counterfactuals, I study the effects of operating the e-hail platform under a stronger network and I waived the booking fee. Beside matching technology and a larger driver network, another advantage ridesharing services have over taxis is that they are subject to less restrictions on fare rates and shift schedules. Currently, TLC is experimenting on two pilot programs which include the flexible fare and leasing pilots. The flexible fare program attempts to allow yellow and green taxis to offer upfront, binding fare quotes to passengers using e-Hail apps. The flexible leasing program relaxes the lease cap and shift schedule rules, allowing participant lessors to set leased shift length less than 12 hours. Previously, TLC rules required that lessors lease Taxicabs through shifts of 12 consecutive hours and thereby prohibit leases for shorter periods of time. To maximize the lease payments, most medallion owners set the shift changing times around 5 pm. When the shift changes, day shift drivers return to the garage and prepare the vehicles for the next shift, resulting in a supply decrease at rush hour. The goal of the flexible leasing program is to increase the number of vehicles on road and serve more passengers during the evening rush hours.

I present a spatial search and match model in which pure street hail and e-hail cab drivers compete against each other to maximize their expected earnings. Pure street hail drivers only pick up waiting passengers on the street. E-hail drivers pick up street hail rides and also respond to requests from e-hail platforms. In each time period, demand for street hail and e-hail taxicabs arrive according to an exogenous process, and drivers make rational spatial search decisions as they cruise around New York City searching for passengers. A spatial competitive equilibrium under this model is defined as a set of beliefs about street hail and e-hail pickup probabilities which are consistent with the spatial distribution of vacant cabs induced by aggregated strategic
search behavior of drivers. In the model, cab drivers take net demand, passengers who chose to ride taxicab, as exogenous.

To solve for competitive equilibria in the model, I introduce an iterative belief update algorithm that utilizes the contraction property of the smooth max operator and numerically converges at a fast rate. In each iteration, I allow players to update their beliefs according to the realized states of the previous iteration. For a given matching function parameter vector, the beliefs converge within 10 iterations. To incorporate demand side response to regime changes, I model passengers’ choice of transportation with a nested logit utility structure to estimate passengers’ price and travel time elasticities. Passengers with deterministic origin-destination location pairs choose their transportation modes among subway, taxicabs, and ridesharing services (e.g. UberX). The challenge of demand estimation for taxis arises from the fact that I only observe passengers when they were picked up by taxicabs. Variables including the number of waiting passengers and the destinations for UberX ridesharing and subway metro passengers are not observed. I infer the posterior distributions of passengers’ destinations and develop a multiple imputation estimator for a discrete choice model of transportation choice between taxi, UberX and subway metro\(^1\). In each imputation an individual-level dataset is constructed which includes the imputed total travel time and fare for the actual chosen and hypothetical equivalent trip had the passenger chosen other alternatives. Real time urge price multipliers and waiting times data were collected through the official Uber application programming interface. For taxicabs, waiting times are unobserved but I use a waiting time approximation model built on the one presented in Frechette, Lizzeri and Salz (2017) and define a “heatmap rule”

car movement within a location.

\(^1\)including the time to walk to the closest stations
Taxicabs and Uber compete on price and waiting time dimensions. Under the computed equilibrium, taxicabs offers a shorter waiting time but higher net-of-surge fare than UberX in Manhattan, and a longer waiting time in other boroughs, whereas UberX offers a lower fare on average. This paper uses datasets collected from various sources, including TLC, Uber, and Google Maps. To my knowledge, this is the first paper to analyze the e-hail taxi request and response dataset recorded by TLC. On the other hand, real time price and waiting time data for UberX collection from Uber through its application programming interface (API) was also featured in Bian (2018) and Shapiro (2018), but the data on UberTAXI is used for the first time in this paper. The model fits the data well.

The model is estimated by the method of simulated moments by finding parameters that best match the number of pickups in data. Spatial distributions of vacant cabs are consistent with Lagos (2000): most vacant cabs search in midtown and downtown Manhattan. Under estimated parameters, the model matches unmatched moments on shift earnings and drivers’ proportion of shift time spent on trips (utilization rate thereafter). The model estimates imply that earnings per hour are $26.32 and $27.58 for pure street hail and e-hail drivers respectively, whose utilization rates are 48.93% and 49.13%. The e-hail matching parameter is estimated at 5.36, implying that the average e-hail dispatch is able to reach 5.36 drivers partners on the network for each request. On top of that, TLC mandates that e-hail matches must be made within 0.5 miles in downtown Manhattan and 1.5 miles in other area and this further decreases the likelihood of request completion. In the data, only 60% of e-hail requests are completed by a nearby e-hail cab. There is a net gain of using e-hail in terms of higher revenue and better utilization of drivers’ shift times.² The magnitude

²E-hail costs paid to dispatch platforms by medallion owners are proprietary and are not modeled.
of the gains is low due to low arrival rate of e-hail requests and the small network size. On the demand side, the willingness to pay for each reduced minute of travel time is estimated at $0.536.

The following counterfactual experiments were studied. First I exogenously replaced a fixed percentage of street hail demand with e-hail demand and determined its effect on drivers’ incomes and utilization rates. I estimate that replacing 30% street hail demand with e-hail technology generates 22% earnings and 23% utilization rate gains for e-hail drivers, and this premium decreases when more drivers employ the technology. It is worth noting that if the substitution pattern is uniform across all locations, vacant yellow cabs are further disincentivized to search outside Manhattan. In the next set of counterfactuals I allow the network strength of e-hail programs to match that of UberX and waive the $2 booking fee. I found that the waiting time for taxi is reduced and becomes lower than UberX in most places, and consumer surplus is $0.15 for each commuter traveling in New York City. Finally, allowing drivers to work in flexible schedules increases the number of active cabs on the road during the evening peak (4 p.m. to 8 p.m. window), causing the number of matches to increase by 3,710 and average shift earnings to increase by $7.2. In the appendix, I also explored the efficiency gains of a carpool matching algorithm in which passengers going to the same destinations can share a taxi ride. Under the condition of the first set of counterfactuals, carpooling increases the number of matches produced by around 5%.

In light of the public discussion on the lack of regulation of ridesharing services and the deterioration of taxi drivers’ revenue, counterfactual experiments studied in this paper show that even when ridesharing services remain loosely regulated, when network strength is sufficiently strong and there is no additional cost on passengers’ side, e-hail cabs have the potential to generate efficiency and revenue gains for drivers while
improving consumer welfare. This suggests that a city-operated, universal matching matching platform without surcharges can be a way to improve the taxicab industry and reclaim lost market share.

1.1.1 Related Literature

The empirical literature on NYC taxicab market flourished with the release of TPEP detailed taxicab trip records in 2009. The trip time and location information enabled researchers to study supply side behavior directly. For example, Farber (2014) utilized trip record data from 2009 to 2013 to estimate the wage elasticities of taxi drivers’ labor supply. He obtained positive estimates and found that drivers that remain in the market learn more quickly than those who exited the market. This is echoed by Haggag (2017) who found evidence of learning by doing among newer drivers in their first 20 shifts: drivers accumulate neighborhood-specific experience and improve their earnings by making optimal search decisions.

Several seminal empirical works studied endogenous search frictions in the NYC taxicab market, modeled as dynamic search and match game. To study the potential impact of a centralized dispatch platform and relaxed medallion ownership restrictions, Frechette, Lizzeri and Salz (2017) (FLS thereafter) modeled a dynamic game in which cab drivers play against each other in Manhattan and make intertemporal decisions as to when to enter and exit the market within their designated shift times. While FLS modeled search friction resulting from intertemporal mismatch of demand surge of shift changing times, it did not consider trip demand heterogeneity across locations. I model drivers’ spatial search decisions across the entire map of NYC to capture endogenous search frictions. In modeling the effect of digitized matching completed through e-hail dispatch platforms, FLS assumed that e-hail cab drivers only take requests from the platform, abstracting from the possibility that e-hail
cab drivers also pick up passengers on the street. When this is taken into account I find that e-hail matching technology generates improvements in e-hail drivers’ utilization rate and earnings despite the weak e-hail network strength caused by the current oligopolistic e-hail taxi dispatcher market. This is consistent with TLC’s e-hail report, whereas FLS predicted a market segmentation decreases earnings of e-hail drivers when they only pick up passengers through e-hail platforms.\(^3\) Even though in my model entry and exit of cab drivers are taken as exogenous, I address the idea of Frechette, Lizzeri and Salz (2017) that search friction arises from intertemporal mismatch of supply and demand by considering the counterfactual experiment of introducing the TLC flexible leasing pilot program, including more flexible shift schedules and shift sharing. \(^4\)

Buchholz (2018) modeled a dynamic spatial search and match model in the NYC taxicab market for day shift drivers. Despite differences in matching functions and solution methods, his use of locational heterogeneity in passenger arrival rates result in equilibrium spatial distribution of cabs in profitable zones is parallel to the dynamic spatial matching game model presented in this paper. I extend the spatial competition concept in Buchholz (2018) by modeling explicitly passengers’ elasticities on waiting time and introducing peer-to-peer ride-sharing services as a competing alternative. Counterfactual experiments in Buchholz (2018) studied the impact on consumer surplus when taxi fare is allowed to vary by time of the day, pickup location or trip distance, as well as when matching between passengers and taxicabs is centralized on a dispatch platform. When substitution from Uber and metro riders are factored in, I found an overall net gain for consumers, in contrast to the conclusion in Buchholz.

\(^3\)The improvement on earnings is consistent with TLC’s survey published in the e-hail report in http://www.nyc.gov/html/tlc/downloads/pdf/ehail_q1_report_final.pdf

\(^4\)These regimes are being tested by TLC and the Lacus mobile application. See: http://www.nyc.gov/html/tlc/html/industry/lease_pilot.shtml
(2018) that dispatch platform matching generates overall loss for consumers by further drawing cabs to more profitable areas. The model presented in this paper finds overall gains for on the demand side, as will be discussed in section 5.

With the publicly available application programming interface (API) that grants access to Uber’s real time database, two seminal papers built directly on Buchholz’s (2018) model to study the competition between peer-to-peer ridesharing applications and taxicabs. Bian (2018) included Uber drivers in a spatial equilibrium similar to Buchholz (2018) and characterized the benefits of Uber as the improvement on matching efficiency by network effects. Shapiro (2018) studied the gains in consumer surplus across different locations and found that the welfare improvement with the addition of Uber is greater in locations with small number of potential consumers per unit of area. In this paper, I take the market with ridesharing services as the status quo and study the welfare effects of various regulation policies and technological innovations in the taxicab market. My methodology differentiates from the above works: Bian (2018) estimated and measured network effect in terms of passengers’ mean utility, whereas I model the benefits from a dispatch network through variations in passengers’ waiting times and the equivalent willingness-to-pay. It is acknowledged that the variables and information revealed in the Uber API data set is inadequate to model robust Uber drivers’ behavior. In particular, ridesharing drivers do not operate on regular shift hours that apply to cab drivers and their labor decision respond real time demand and surge pricing (Chen and Sheldon (2015)). For instance, it is interesting to note that Bian (2018) assumes 3,000 Uber vehicles operated in New York City (that seems to be implausible with the observed waiting time and the number of trips) whereas Shapiro (2018) estimated 21,356 Uber vehicles, and both papers were able to impose assumptions and estimate structural parameters to rationalize the market shares, suggesting that identification of Uber supply and demand and there-
fore counterfactuals involving these variables are sensitive to model assumptions. In this regard, model estimates face less challenge in robustness, as the focus and research question of this essay lie on identifying agents’ behavior and welfare in the taxicab market (instead of the entire for-hire vehicle industry) under various regimes.

Recently, there is a literature on Uber’s welfare effects on the labor market. Chen and Sheldon (2015) used panel trip record data obtained from Uber Technologies and found evidence that surge pricing significantly increased supply of rides on Uber platform when there were price surges. Other studies that made use of the internal panel trip record for Uber rides have found that Uber’s flexible shift schedule generated improvements in labor surplus (in excess or reservation wage, when compared to the traditional taxicab fixed shift system) (Chen et al. (2017)), higher utilization rates (Cramer and Krueger (2016), Hall, Horton and Knoepfle (2018)) and provided opportunities for driver partners to smooth fluctuations in their income (Hall and Krueger (2015)). On passengers’ side, studies have made use of discrete choice models (Lam and Liu (2017)) and regression discontinuity (Cohen et al. (2016)) to estimate the improvement in consumer surplus generated by Uber’s entry.

A number of seminal works in the information engineering and operations research literature studied possible improvement schemes of the taxicab industry under technical and algorithmic approaches. Examples of such schemes include artificial intelligence routing system to improve drivers’ search decisions (Wang (2014)), e-hail platform with dynamic schedule matching (Bai et al. (2014)) and trip sharing (Santi et al. (2018)). I borrowed a few elements from these papers, including the timing of taxicab movement under a discretized time space in Wang (2014) as well as the matching function for e-hail presented in this paper (Bai et al. (2014)), in which the dispatch platform generates ordered lists that assign rides.
To my knowledge, this is the first work to evaluate the impact of a flexible leasing pilot program on the taxicab market. In terms of methodology, I developed an iterative method to solve for equilibrium in a class of oligopoly games in which agents’ choice probabilities are expressed in the logit (softmax) form. The iterative method that maps beliefs on a parameter to the realized parameter implied by players’ strategies utilizes the fact that the softmax function is 1-Lipschitz, and under plausible differentiability and boundedness assumptions, successive iteration on players’ beliefs converges to a fixed point that corresponds to a self-fulfilling belief search equilibrium. This provides an algorithm to solve for oblivious equilibria, a concept introduced in Benkard, Van Roy and Weintraub (2009) under a class of oligopoly games with many players whose state is a parameter or a function of the distribution of other players. An alternative interpretation of my belief iteration algorithm is it models a reinforcement learning process where agents play the game repeatedly and each time update their beliefs using the payoffs and revealed information on the realized parameter of distribution. This is consistent with the empirical findings of Haggag (2017) which found learning by driving among taxi drivers. A detailed analytic discussion is included in the Appendix D. It can be shown that this iterative method is a variation of Gauss-Seidal successive iteration and it can be shown that the method converges rapidly.

The organization of the paper is as follows: section 2 describes the format and sample construction. Section 3 outlines the model, equilibrium definition and solution algorithm. Section 4 describes identification and estimation procedures. Section 5 presents the estimation results. Section 6 covers the counterfactual experiments and section 7 concludes. Technical topics including an analytic discussion of the equilibrium solution algorithm, details on the multiple imputation procedure used in demand
estimation and a description of taxicab waiting time models in Manhattan and other
boroughs are covered in the appendices.

1.2 DATA DESCRIPTION

To obtain supply and demand variables for taxicab industry, three datasets were col-
lected from TLC and Uber. Trip record data for yellow and green cabs as well as
pickup information for all ridesharing trips were collected by TLC, e-hail requests
response data from June 2014 to the end of 2016 and real time surge pricing and
waiting time data were collected from Uber using its application programming inter-
face (API). The subsections below describe each data set in detail.

1.2.1 TAXICAB AND LIMOUSINE COMMISSION TRIP RECORD DATA

Taxicab and Limousine Commission (TLC) enacted that Technology Passenger
Enhancement Project (T-PEP), which mandates that all taxicabs install devices that
enable automated trip data collection in 2009. The trip record data set includes all
trips completed by licensed taxicabs (yellow and green). Each observation includes:
pickup and dropoff date and time, geo-coordinates of pickup and dropoff locations,
trip distance, payment items incurred in the trip (fare, tax, toll, tips and extra
surcharges), a fare rate code that dictates how the fare was calculated. Trip record
data for yellow and green cab trips that took place from Monday to Thursday in
June 2016 are used for model estimation. I chose data in June 2016 to control for
heterogeneity in demand pattern and to make computation feasible, and because it
has the most overlapping taxicab and Uber observations.

In June 2013 New York City issued boro cab licenses which permit owners of Street
hail livery vehicles (also known as green cabs) to operate and pick up passengers
outside the yellow zone and the airports. As shown in the Fig. 1.1, the “yellow zone” refers to the area of Manhattan below West 110 Street and East 96 Street, which is the restrictive zone in which only yellow cabs can pick up passengers. Color intensity for heat map in the same figure represents pickup densities for yellow cabs. As illustrated, yellow cabs are most active in the yellow zone.

![Map of Manhattan showing yellow zone and heat map]

**Figure 1.1:** Yellow zone and heat map of passenger pick ups.

<table>
<thead>
<tr>
<th></th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Earnings</td>
<td>16.92</td>
</tr>
<tr>
<td>Distance (miles)</td>
<td>3.02</td>
</tr>
<tr>
<td>Duration (minutes)</td>
<td>17.30</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
</tr>
</tbody>
</table>

*Exceptions are made if airport trips are pre-arranged.*
Table 1.2: Descriptive Statistics for green cab trips.

<table>
<thead>
<tr>
<th></th>
<th>green mean</th>
<th>median</th>
<th>5th Pctl</th>
<th>95th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>trip earnings</td>
<td>15.11</td>
<td>11.62</td>
<td>5.80</td>
<td>36.30</td>
</tr>
<tr>
<td>distance (miles)</td>
<td>2.84</td>
<td>1.85</td>
<td>0.52</td>
<td>8.40</td>
</tr>
<tr>
<td>duration (minutes)</td>
<td>23.21</td>
<td>10.92</td>
<td>2.18</td>
<td>37.14</td>
</tr>
<tr>
<td>observations</td>
<td></td>
<td></td>
<td></td>
<td>729,755</td>
</tr>
</tbody>
</table>

The graphs below show the distribution of shift starting time and the number of pickups at each given time in a day. Most day shift drivers started their shift at around 6 am and night shift drivers starting time peaks between 5:30 pm and 6 pm.\(^6\) Because day shift cab drivers need to head back to the garages and hand the car to night shift drivers, there is a drop in number of active cabs during 5 to 6 pm. The mismatch of demand surge and taxicab supply is a source of search frictions in the matching market, and this is known as the “witching hour”\(^7\) phenomenon among New York city residents.

\(^6\)As discussed by FLS, the earnings of the medallion owner is maximized when he sets the shift change time at around 5pm. The implication of the estimated value function is consistent with this fact, too. See section 4.

\(^7\)https://www.nytimes.com/2011/01/12/nyregion/12taxi.html
Figure 1.2: Distribution of shift start time.

Figure 1.3: Average daily number of pickups for yellow (bold) and green cabs (dotted).

TLC mandates that for-hire vehicle (FHV) dispatchers must submit trip data. The FHV trip record data consists of all pickup time and location (in terms of TLC
taxi zones) for all licensed for-hire vehicles, including the four operating peer-to-peer ridesharing applications in New York: Uber, Lyft, Via and Juno. Among its competitors, Uber has the largest market share, making up 50% of all FHV rides. Other FHV pickups refer to traditional pre-arranged vehicle services, such as black cars and luxury limousine.

**Table 1.3:** Number of pickups for each for-hire vehicle in June 2016.

<table>
<thead>
<tr>
<th></th>
<th>Uber</th>
<th>Lyft</th>
<th>Via</th>
<th>Juno</th>
<th>Other FHV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5,358,145</td>
<td>1,074,722</td>
<td>517,964</td>
<td>332,933</td>
<td>3,586,492</td>
</tr>
</tbody>
</table>

Peer-to-peer ridesharing apps such as Uber, Lyft and Via are regulated by TLC. Drivers on these platforms must apply for licenses issued by TLC, but matching mechanisms and prices are subject to little regulations. Most ridesharing applications are operated through smart phone applications which serve as centralized matching platforms. Compared to tradition cab dispatchers, these platforms have access to much larger networks of customers and drivers, and they are automated to match cars and passengers in any volume and time. Market share of taxicabs has consistently declined. Since the end of 2016, the market share of taxis is falling behind of all ridesharing apps combined. For passengers, app-based FHVs complements taxicab in the following following senses. First, without the shift change restriction cab drivers face during the evening demand surge around 5 to 6 pm, Uber and Lyft pickups do not exhibit the decreasing pattern in pickups that taxicabs experience, thereby reducing one source of mismatch in the market. Moreover, dynamic pricing schemes employed by these app dispatchers are designed to induce drivers’ entry decisions and help serve passengers when there are demand surges (Chen & Sheldon (2015)).

---

8The ridesharing driver license system is mainly of administrative purpose and it is different from the medallion system as it is non-transferable.
shares of combined app-based FHV in each location. Zones colored in dark grey are locations at which number of app-based pickups exceeded the number of taxicab pickups. While most of the taxicab pickups took place in Manhattan, app-based FHV served passengers from other boroughs as well, establishing a large market share in these zones. Taxicab maintained its market share in Manhattan, north Brooklyn and JFK airport only.

**Figure 1.4:** Taxi and app for-hire vehicle pickups.
Figure 1.5: Market shares of taxicab and other for-hire vehicle.

Two challenges arise from the data structure of the TLC trip record. First, total number of waiting passengers and searching cabs are not observed, and a pair of cab and passenger is observed only when they are matched. In other words, the spatial distribution of cabs that failed at passenger searching and passengers that waited and left without a match are unobserved. This creates a challenge in identifying the underlying mechanisms of data generating process, in particular, the matching function and variation of passenger waiting times. Buchholz (2018) imposes the assumption that the matching between waiting passengers and searching vacant cabs follows the urn-ball matching function, which allows identification of demand and matching parameters from the trip record data. FLS exploits geographical characteristics of Manhattan and estimate the matching function through grid point simulations. Brancaccio, Kalouptsidi and Papageorgiou (2018) studies matching between bulk ships and exporters in the ocean transportation services market where only matches and ships are observed and it presents a non-parametric method of matching function estima-
tion imposing assumptions including monotonicity and constant returns to scale. I employed a parametric assumption on the matching function with microfoundation consistent with spatial heterogeneous grid point simulations built on FLS. This will be discussed more in Appendix F. The second challenge arises from the lack of panel data structure, making identification of drivers’ attributes such as searching times, joint distribution of entry and exit, and breaks difficult. Ideally, the data contain GPS trajectories of each cab, as in the case of electrical engineering literature on the taxi industry in China (Chen et al. (2015) and Jiang and Zhang (2018)). On the other hand, Farber (2005) used data consisting of driving logs provided by a minifleet company, where he observed drivers’ rest durations and timestamps. It is noted that TLC trip record data was available in panel data format for year 2013, and an equilibrium was estimated using trip record data of 2013 (June) to check the model’s ability to fit unmatched moments as such drivers’ searching times and utilization rates.

1.2.2 E-hail Trip Data

The taxicab industry attempted to revolutionize matching mechanisms too, as TLC approved several cab hailing phone applications (referred as “e-hail programs” in accordance to section 52-27(a) of TLC rule) in December 2012. e-hail programs must meet a set of TLC rules, notable requirements include: 1. apps may not charge a fee on the passenger unless the driver receives the full fare amount. 2. participant apps cannot require a passenger to pay a tip to the Driver. 3. requests transmitted to a driver must not reveal the passenger’s desired destination. 4. the matched driver must be within 0.5 miles from the passenger for requests originating from south of 59th street, and 1.5 miles elsewhere. 5. passengers may not make e-hail request at the airports.
I requested the trip data of all e-hail taxicab rides from June 2013 to December 2016 from TLC under the freedom of information law. The data set include all taxicab ride requests from approved e-hail apps. They include phone applications (apps) that are designed specifically for hailing cabs, such as Hailo and Curb, and ride sharing apps that support cab hailing services, such as the UberTAXI option under Uber. The data set contains the datetimes, locations and request outcomes for all requests made on e-hail apps.

Table 1.4: Outcome of e-hail requests.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>2013 (Jun to Dec)</th>
<th>2014</th>
<th>2015–2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancelled by driver</td>
<td>41,301</td>
<td>154,758</td>
<td>391,183</td>
</tr>
<tr>
<td>Cancelled by passenger</td>
<td>145,044</td>
<td>507,729</td>
<td>1,251,154</td>
</tr>
<tr>
<td>Completed by green cabs</td>
<td>37,722</td>
<td>1,232,676</td>
<td>3,540,561</td>
</tr>
<tr>
<td>Completed by yellow cabs</td>
<td>247,714</td>
<td>706,669</td>
<td></td>
</tr>
<tr>
<td>No response</td>
<td>404,211</td>
<td>799,867</td>
<td>1,014,333</td>
</tr>
<tr>
<td>Total requests</td>
<td>875,992</td>
<td>3,415,592</td>
<td>6,197,232</td>
</tr>
</tbody>
</table>

Each e-hail request has three possible outcomes: cancelled, completed or no response. Request flagged as no response are requests never accepted because there were no cabs nearby, or no cabs that were matched accepted the request. The number of green cabs grew significantly in 2014 and green cabs became the major type of vehicles that complete request by the end of 2014. The number of drivers that were equipped with e-hail programs are not revealed in the data, but are given in the e-hail pilot program quarterly reports published by TLC. At the end of year 2013, there were 5,060 drivers that completed at least one e-hail trip. The number grew to 8,407 by the end of year 2014.9

9TLC does not reveal how many of these drivers are yellow and green cab drivers.
Figure 1.6: Total (solid) and completed (dotted) e-hail requests in the year 2016.

E-hail request completion rate was 32.61% in 2013 and grew to 57.13% in 2016. When a trip was completed, the data reveals whether the request was completed by a yellow cab, green cab or licensed for-hire vehicle for the e-hail record prior to 2015.

Table 1.5: Location of requests completed by yellow cabs by borough, 2013-2014.

<table>
<thead>
<tr>
<th>Borough</th>
<th>Yellow cabs</th>
<th>Requests</th>
<th>Completion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manhattan</td>
<td>319,464</td>
<td>1,113,748</td>
<td>28.68%</td>
</tr>
<tr>
<td>Brooklyn</td>
<td>403,573</td>
<td>1,647,622</td>
<td>24.49%</td>
</tr>
<tr>
<td>Queens</td>
<td>81,201</td>
<td>332,504</td>
<td>24.42%</td>
</tr>
<tr>
<td>Bronx and Staten Island</td>
<td>2,040</td>
<td>24,195</td>
<td>8.43%</td>
</tr>
<tr>
<td>Total</td>
<td>806,278</td>
<td>3,118,070</td>
<td>25.86%</td>
</tr>
</tbody>
</table>
1.2.3 Real-Time Uber Pricing and Waiting Time Data

Real-time data on surge pricing and waiting times was obtained from Uber database through its official application programming interface (API). I programmed automated queries on real-time pricing and estimated time of arrival at the centroid of each TLC taxi zone (see Fig. 1.1). Data was continuously collected starting from Feb 19 to Sep 9, 2016. The median time difference between two queries at the same zone is 11.5 minutes. Each query resulted in a response by Uber database including what the surge multiplier and estimated waiting time (ETA) would have been for a ride request originating at the centroid of the zone been requested at the time of query. The waiting time estimates provided by Uber were rounded to the nearest minute. Surge multiplier is a factor that is multiplied to base fare, set by Uber. The minimum value of the surge multiplier is 1. In other words, final fare paid by passengers is \( \text{surge multiplier} \times (\text{price per mile} \times \text{distance} + \text{price per minute} \times \text{duration}) \).
Surge multipliers (unique with respect to pickup locations) are updated frequently and responds instantaneously to real time supply and demand. Uber provides peer-to-peer ridesharing services under several options. The surge multiplier applies to UberX trips, at the time of data collection, UberPOOL fares were calculated upfront and the surge multiplier for UberPOOL requests was set to 1.

Table 1.6: Descriptive statistics for API data.

<table>
<thead>
<tr>
<th></th>
<th>Manhattan</th>
<th>All zones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of surge multiplier</td>
<td>1.0641</td>
<td>1.0255</td>
</tr>
<tr>
<td>Median of surge multiplier</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean of waiting time (minutes)</td>
<td>3.907</td>
<td>5.158</td>
</tr>
<tr>
<td>Median of waiting time (minutes)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Observations</td>
<td>739,264</td>
<td>3,023,289</td>
</tr>
</tbody>
</table>

According to Uber, surge pricing schedules are adopted to ensure passengers with high willingness to pay get matched when there is excess demand.\footnote{https://help.uber.com/h/e9375d5e-917b-4bc5-8142-23b89a440eec} The data reveals that surge multipliers were indeed higher during morning (8 am to 10 am) and evening peaks (4 pm to 7 pm), and at each time in a week they were higher on average in Manhattan than other boroughs. Also, values of surge multipliers were positively correlated with waiting times, suggesting that the surge pricing tends to apply when markets were tight.
Besides peer-to-peer ridesharing, Uber also partnered with taxicab drivers and provided e-hail services to passengers with the UberTAXI option. UberTAXI matched partner cabs with passengers with a $2 booking fee. Unlike other Uber services, passengers were required to pay the trip fare to cab drivers directly, and the fare amount was calculated with the standard NYC meter fare. UberTAXI waiting time queries only returned values when there were partner cabs around the request locations, and
there were 711,820 observations. UberTAXI has been quietly shut down and become unavailable since August 24, 2016.\textsuperscript{11}

1.3 The Model

It is observed from various data that demand was highest in Manhattan, as the number of taxicab matches (resulted from street hail) produced in Manhattan were higher than other boroughs (Fig. 1.1) and the mean surge multipliers were higher. Despite e-hail demand is highest in boroughs outside Manhattan, the e-hail demand intensity from outer boroughs did not incentivize yellow cabs to search for rides in the outer boroughs, reflected by the low e-hail completion rates. In this section, I present a spatial dynamic search and model of yellow cab drivers. In particular, I account for drivers who have access to e-hail matching technologies in order to study how e-hail impacts drivers search decisions under equilibria in data and counterfactual setups. Matching function parameters in the model estimated in next section describe the matching efficiencies and net demand across each time period and location.

1.3.1 Model Environment

Time and location spaces are discretized in the model. Each day is partitioned into 144 10-minute periods. Each yellow cab driver\textsuperscript{12} starts his shift with length $\tau_0$ at time period $t_0$ and zone $z_0$, which are assigned to him according to an exogenous distribution $F(t_0, \tau_0)$. The map of New York City is grouped into 20 zone partition where cab drivers cruise and search for passengers.

\textsuperscript{11}No memorandums were given as to why the service was no longer available.

\textsuperscript{12}I model endogenous decisions of yellow cab drivers only, and I assume an exogenous number of green cab drivers search in the zone out of Manhattan. Because the model seeks to study cab drivers’ problem under an exogenously given net demand, ride sharing drivers (e.g. Uber and Lyft) are not accounted.
There are two types of drivers in the model: pure street hail cab drivers only pick up passengers who hail cabs on the streets. E-hail cab drivers are equipped with phone applications with centralized dispatch platforms and when they are matched with e-hail requests, e-hail drivers can choose to accept or decline them. They can also pick up passengers on the streets hailing for cabs. The e-hail usage rates among yellow and green cab drivers are exogenous, and drivers do not switch types. There is an exogenous number of green cabs searching outside the yellow zone in Manhattan, whom yellow cab drivers at the same locations compete with. All green cabs search exclusively in the non-Manhattan location zone and a fixed proportion of green cabs are equipped with e-hail applications.\(^{13}\)

At the beginning of each time period \(t\), the net demand for taxi - passengers who want a taxi ride arrive according to an stochastic process, which is considered \textit{ex post} exogenous from drivers’ viewpoint as they do not consider feedback of passengers’ behavior caused by their search decisions. Drivers develop a system of the beliefs of probabilities of street hail and e-hail pickups, \(\{P(z,t), P^e(z,t)\}\). Conditional for the location and time location pair \((z,t)\), the action space of each vacant cab driver consists of his search decision whether to cruise and search in zones that are accessible to the current zone, to stay in the current location, or to take a break for the coming period.\(^{14}\) The actualized probability of pickup is a function of the cab distribution resulted from the spatial search decisions of all vacant cab drivers. For a constant arrival of passengers, a higher number of cabs in a location decreases each driver’s probability of picking up a passenger.

\(^{13}\)In other words, green cabs search decisions are not modeled because green cabs can only pick up passengers in only one zone in the model zone partition. Therefore, by default they can only choose to search in that particular zone.

\(^{14}\)Boroughs beside Manhattan are accessible from East Harlem through Robert F Kennedy Bridge, from zones Upper East Side South and Midtown East through the Queensborough Bridge and Tunnel, and from lower Manhattan through Brooklyn and Williamsburg bridges.
When a street hail passenger is picked up, the driver has to start the trip and take the passenger to her destination. One simplifying assumption imposed on drivers is that they will not take street hail trips that end beyond their designated shift times.\textsuperscript{15} While it is not lawful for a cab driver to refuse taking a passenger to the requested destination, this can be viewed as the fact that drivers can infer from his experience and unobserved characteristics whether a hail request will result in a long trip, and decide to take only the ones that are short enough.\textsuperscript{16} On the other hand, TLC mandates that e-hail apps must not reveal information of a e-hail passenger requests beside their originating locations, but an e-hail driver has the right to accept or decline an incoming request. Should an e-hail driver accepts the request near the end of his shift, he faces the risk of finishing the trip after his shift end time and pays the late fee. A trip starts when a cab is matched with a passenger, and the destination and duration follow an exogenous law of motion, and the driver is awarded a fare that corresponds to the transition of location and time states.

By the end of the period $t$, trips, e-hail and street hail matching processes are completed. The subsequent vacant cab distribution at a particular location consists of the unmatched cabs that searched in the area, cabs that just dropped off passengers whose destinations are at that location and cabs that start their shifts net of number of the cab drivers ending their shifts. Unmatched passengers leave and new passengers arrive according to an exogenous stochastic process across locations for period $t + 1$.

\textsuperscript{15}This assumption is justified by the fact that a driver who handed his shift late will be subject to a $30$ fine. Also, endogenizing street hail drivers’ acceptance decisions would greatly increase the computation burden.

\textsuperscript{16}Hailing passengers in New York are said to have gestures that signal drivers their trip distances. For example, a C-shaped finger gesture signals drivers that the passenger will only go for a short distance.
1.3.2 Drivers’ Decision Problem

At the beginning of a 10-minute period \( t \), drivers of vacant cabs make spatial search decisions for the period. Regardless of the drivers’ decision, vacant cabs will stay in the originating zone for the first \( x(z,t) \leq 10 \) minutes of \( t \). Thus, a driver’s decision internalizes the probability of picking up a passenger on their way to the next zone and also at their destination zones.

The model allows for pre-commitment decisions but only when the searching zone is not adjacent and is connected to the current location through bridges or tunnels, where drivers cannot pick up passengers. When the destination zone \( z' \) is not adjacent to a driver’s originating zone, a deterministic time cost to reach the destination \( z' \), depending on \( (z,t,z') \) is incurred and the driver does not re-optimize until he reaches \( z' \).\(^{17}\) For example, a cab that decides to go to Midtown East from Queens will not arrive the Midtown until 1 period (10 minutes) later, and he is not able to pick up passengers during the time spent on Queensborough Bridge on his way.

The state variables for a vacant cab driver’s dynamic decision problem consist of the current (location, time) pair \( (z,t) \) and \( P(z,t), P^e(z,t) \), which encode a driver’s belief about the probability of picking up street hail and e-hail passengers, and the time remaining on his shift. The action space consists of the driver’s choice of searching zone. He can also choose to take a break for a period. The value function is the expected sum of the driver’s fare earnings for his shift.

To sum up, the timing and sequence of possible events for a vacant cab within a period can be summarized as follows:

1. A vacant driver decides to search and cruise to zone \( z' \). The expected time cost associated with his decision is \( \Delta(z, z', t) \geq 1 \)

\(^{17}\)In real life, the time cost may be stochastic. To ease computation burden I assume that drivers condition on their expectation of the time costs.
2. He spends the first \( x(z, t) \) minutes in his originating zone \( z \). He will pick up a passenger if he is matched. When a passenger is picked up, the cab will drop the passenger off at another zone \( z_2 \) at time \( t_2 \) which are drawn from an exogenous transition kernel \( Q \).\(^{18}\) The total payment received by the driver is given by an exogenous payoff function \( \pi(z, t, z_2, t_2) \). In the event that he is not matched with a passenger on his way to \( z' \):

(a) If \( \Delta(z, z', t) = 1 \) (going to an adjacent zone), then he reaches zone \( z' \) and searches for passengers for the remaining \( 10 - x(z, t) \) minutes.

(b) If \( \Delta(z, z', t) > 1 \), then he will not be able to search and pick up passengers until he reaches zone \( z' \) at time period \( t + \Delta(z, z', t) \).

**Pure Street Hail Drivers’ Problem**

Denote drivers’ system of beliefs on the probabilities of picking up a passenger by \( \{P(z, t)\} \). A vacant cab driver chooses his destination zone in the current period to maximize his discounted sum of payoffs. His problem can be represented by the following Bellman equation:

\[
V(z, t) = \underbrace{P(z, t) \sum_{z_2, t_2} Q(z_2, t_2|z, t) \left\{ \pi(z_2, t_2|z, t) \right\}}_{\text{Ride prob on his way}} + \underbrace{V(z_2, t_2)}_{\text{Value at drop off}} + \underbrace{(1 - P(z, t)) \mathbb{E}\left[ \max_{z'} \left\{ W(z, t, z') \right\} \right]}_{\text{No ride in the first } x \text{ minutes, value at destination}}
\]

(1.1)

\( W \) is the action-specific value function when the cab reaches the destination zone \( z(d) \). If the zone is adjacent such that the cab reaches the destination zone, in other

---

\(^{18}\)For each origin-destination pair \((z, z_2)\) at a given time \( t \), the time elapsed is random: \( \exists t_a, t_b \) such that \( Q(z, t, z_2, t_a) > 0 \) and \( Q(z, t, z_2, t_b) > 0 \): there is randomness in trip travel times. I impose deterministic travel time constraint on vacant searching cabs but not trip duration because this randomness does not create a computation burden that requires multiple evaluation of value functions. Also, this helps to capture the randomness due to the rounding errors under the discretized time space.
words, if time cost $\Delta(z, z', t) = 1$, then:\(^{19}\)

$$W(z, t, z') = \begin{cases} P(z', t) \sum_{z_2, t_2} Q(z_2, t_2|z', t) \left\{ \pi(z_2, t_2|z', t) + V(z_2, t_2) \right\} \\ + \left(1 - P(z', t)\right) V(z', t + 1) + \varepsilon(z') \end{cases}$$

(1.2)

In other cases where the time cost is positive and the driver will not arrive the destination zone until $\Delta(z, z', t) > 1$ periods later. The term $W$ becomes:

$$W(z, t, z') = V(z', t + \Delta(z, z', t)) + \varepsilon(z')$$

(1.3)

The term $\varepsilon(z')$ component is the idiosyncratic unobserved utility of going to zone $z'$, this captures components of the driver’s utility unobserved to the econometrician such as traffic conditions and drivers’ heterogeneous preferences for zones.

**E-hail Drivers’ Problem**

Drivers equipped with e-hail applications can both pick up street hail passengers and respond to matching requests from nearby e-hail passengers through centralized matching platforms. When a request is matched with a vacant driver, the driver has the option to accept or decline it. An e-hail driver is knowledgeable of the e-hail matching function and his system of belief on e-hail matching probabilities is $\tilde{P}(z, t)$. He makes two decisions: where to go searching and whether to accept e-hail requests when he receives them. His search decision internalizes the probability of getting e-hail requests. Since he does not observe the destination of the e-hail request and has less signal to infer from (unlike street hail passengers), he faces a risk of going beyond his shift hours and ending his shift late, which he will need pay a penalty.

---

\(^{19}\)As discussed, the set of trip destination time $t_2$ considered lies in the driver’s predeter-

mined shift time, so drivers do not pick up trips that go beyond their shifts.
In the two-stage search setup, I assume that e-hail requests come when the driver arrives a zone. Specifically, at the beginning of time period $t$, a vacant e-hail driver at $z$ receives an e-hail request with probability $P^e(z, t)$. He makes a decision whether to accept it. If he declines the request, he will search for street hail passengers, with pick up probability $P(z, t)$ for the rest of $x$ minutes. If he reached the zone he decided to search for, $z'$, there is a probability $P^e(z', t)$ of receiving an e-hail request. He then decide whether to accept the request or to search for street hail passengers with probability $P(z', t)$. The value function $V$ and action-specific value functions $W$ are:

$$W^e(z, t, z') = P^e(z, t) \mathbb{E}\left[ \max\left\{ W^{e, 1}_1(z, t), W^{e, 1}_0(z, t, z') \right\} \right] + (1 - P^e(z, t)) W^{e, 1}_0(z, t, z') + \varepsilon^e(z')$$  \hspace{1cm} (1.4)

Street hail search, equivalent to decline

$$V^e(z, t) = \max_{z'} \{ W^e(z, t, z') \}$$ \hspace{1cm} (1.5)

The terms $W^{e, 1}_1$ and $W^{e, 1}_0$ are defined as the value of accepting and declining any possible e-hail request from the dispatching platform in the first $x$ minutes in its his search (spent in originating zone $z1$). Let $W^{e, 2}_1$ and $W^{e, 2}_2$ be the value of accepting and declining an e-hail request in the second stage (where applicable). They are given by the following set of equations.

$$W^{e, 1}_1(z, t) = \sum_{z_2, t_2 : t_2 - t \leq \tau} Q^e(z_2, t_2 | z, t) \left( \pi(z_2, t_2 | z, t) + V^e(z_2, t_2) \right)$$

value of accepting trips ending within shift

$$+ \sum_{z_2, t_2 : t_2 - t > \tau} Q^e(z_2, t_2 | z, t) \left( \pi(z_2, t_2 | z, t) - \rho \right) + \varepsilon^{e, 1}_1$$ \hspace{1cm} (1.6)
The terms $W^{e,2}_1$ and $W^{e,2}_0$ are zero if there is a time cost travel to the destination searching zone. When the time cost is zero, the terms are defined as:

$$W^{e,2}_1 = \sum_{z_2,t_2:t_2-t_2 \leq \tau} Q^e(z_2,t_2|z',t)(\pi(z',t,z_2,t_2) + V^e(z',t'))+ \sum_{z_2,t_2:t_2-t_2 > \tau} Q^e(z_2,t_2|z',t)(\pi(z',t,z_2,t_2) - \rho) + \varepsilon^{e,2}_1 \tag{1.8}$$

$$W^{e,2}_0 = P(z',t) \times \sum Q \cdot (\pi + V^e) + (1 - P(z',t)) \beta V^e(z',t+1) + \varepsilon^{e,2}_0 \tag{1.9}$$

I assume unobserved utility terms follow an extreme value 1 distribution with scale parameter $\sigma$. So the expected maximum terms $\mathbb{E}[\max_{z'} \{ W(z,t,z') \}]$ can be rewritten as $\sigma \log \left( \sum_{z'} \exp \left( \frac{W(z,t,z')}{\sigma} \right) \right)$.

Green cab drivers are accounted for in the model. The percentage of e-hail drivers is the same for yellow and green cabs. Green cabs are restricted to pick up passengers in non-Manhattan and non-airport areas. Since these areas aggregate to a single zone in the model that green cab drivers are constrained in, I treat them as exogenous agents that participate in the matching process in non-Manhattan zones.
1.3.3 Equilibrium

Under the model environment, cab drivers play a dynamic stochastic spatial matching game in the New York City taxicab market. Because of the large number of players, I employ a mean-field approximation equilibrium concept that reduces the computational complexity. This mean-field equilibrium known as oblivious equilibrium introduced in Weintraub et al. (2008). Drivers hold beliefs of pickup probabilities as a proxy of equilibrium strategies of all players in the game. In equilibrium the optimal strategies are consistent with the rational expectations of the belief probabilities. I present the formal definition as follows:

**Definition:** A competitive equilibrium in this model is \( \{ S_0, V, V^e, \phi, \phi^e, \{ P(z,t), P^e(z,t) \}_{z,t} \} \):

- an initial state distribution \( S_0 \), value functions and decision rule pairs \( V(z,t;P) \), \( \phi(z'|z,t;P) \) and \( V^e(z,t;P,P^e) \), \( \phi^e(z'|z,t;P,P^e) \) and systems of beliefs on street hail and e-hail probabilities \( \{ P(z,t), P^e(z,t) \}_{z,t} \) such that:
  - \( V, V^e \) and \( \phi, \phi^e \) solve the drivers’ optimization problem.
  - The actualized systems of street hail and e-hail pickup probabilities induced by \( (S_0, \phi, \phi^e) \) are consistent with drivers’ systems of beliefs \( \{ P(z,t), P^e(z,t) \}_{z,t} \).

In equilibrium, drivers maximize their discounted sum of payoffs from their shifts based on their beliefs on the probability of getting matched, internalizing the locations and strategies of other drivers competing for rides. Furthermore, their beliefs are correct and self-fulfilling. This implies that their search strategies will aggregate to a spatial distribution of vacant cabs that induces passenger pickup probabilities that are consistent with the beliefs they hold. In real life, the ability to correctly update beliefs can be attributed to sharing of information among drivers from the same dispatch base or online community, combined with a gradual accumulation of experience of drivers,
or “learning by doing”, as discussed in Haggag et al (2017). In fact, the following section introduces an algorithm which converges to equilibrium when drivers continuously update their beliefs after each episode of the matching game.

**Solving for Equilibrium**

Let \( \Gamma \) be the operator that maps a system of beliefs on pickup probabilities \((P, P^e)\) to the actualized pickup probabilities \((P, P^e)\) induced by the optimal strategies given the beliefs and initial state \(S_0\).Suppressing notation on \(S_0\), equilibrium exists if:

\[
[P, P^e] = \Gamma(P, P^e)
\]

In other words, an equilibrium system of belief must be a fixed point of \( \Gamma \). Because probability vectors must be on the closed bounded set of \([0, 1]^N\), \( \Gamma : [0, 1]^N \rightarrow [0, 1]^N \) is a function that maps onto itself. Here \( N = 2 \times 144 \times 20 = 5,760 \) is the number of pickup probabilities because for each of street hail and e-hail, there are 144 time periods and 20 locations pickup probabilities that drivers hold beliefs on. I argue here without an analytic proof that the \( \Gamma \) is continuous in \([P, P^e]\): let \( m(C) \) be the aggregate matching function condition on the exogenous demand, then \( \Gamma(P, P^e) = \frac{m(C(P, P^e))}{C(P, P^e)} \) with \( C(P, P^e) \) the spatial distribution of cabs induced by \((S_0, \phi(P), \phi^e(P, P^e))\). As the terms \( P(z, t), P^e(z, t) \) are linear in the action-specific value functions, it follows that the action-specific value functions \( W, W^e \) are continuous in the belief vector. Also, because the unobserved utility terms are extreme value distributed, it follows that optimal strategies \( \phi, \phi^e \) are soft max functions of the action-specific value functions. The implied spatial distribution of cabs due to shift start and end times, trip pickup and dropoff transition probabilities and strategies is a system of accounting equations in \([\phi, \phi^e](z, t) \cdot C(z, t)\). Therefore \( \Gamma \) is continuous in \([P, P^e]\) as long as the matching
function $m$ is continuous. An intuitive explanation of the continuity of $\Gamma$ is that if the matching process is smooth, small perturbations of beliefs should not induce abrupt changes in the actualized pickup probabilities. A more analytic representation of the operator $\Gamma$ and proof of its continuity is included in Appendix D.

By Brouwer fixed point theorem, since $\Gamma$ is continuous in the belief vector $[P, P^e]$, a fixed point exists for $\Gamma$. Given the existence of equilibrium, it can be solved using a successive iteration over relaxation method for a fixed point, using the fact that $\Gamma(p)$ and its derivative are bounded and $\alpha \Gamma(p) + (1 - \alpha)p$ converges to a fixed point for some $\alpha < 1$. Because the evaluation of $\Gamma(P, P^e)$ involves the realization of cab distribution through laws of motion for shift schedules and trip transitions which is difficult to derive the analytic form, I use simulated values to approximate the function value. Conditional on a realized value of initial states $S_0$ drawn from the estimated shift starting time, zone and shift length distribution, the trajectory of each cab can be tracked with the simulated decisions and matches from decision rules $[\phi, \phi^e]$ and actualized pickup probabilities $\Gamma(P, P^e)$.  

\footnote{A discussion on the convergence criterion for this successive belief iterative method is included in Appendix D.}
Algorithm 1 Simulated fixed Point iteration over beliefs

1: Initialize $n = 0$
2: Start with an initial system of belief $P^{(n)}$, $P^{e(n)}$ and solve for dynamic programming optimal policies $\phi^{(n)}$, $\phi^{e(n)}$
3: while non-convergence of beliefs do
4: Simulate $K$ paths of cab distributions $\{C(z,t)\}$ from the estimated shift starting time, zone and shift length distribution $S_0$:
   - For each $t$, simulate decisions by each cab from $[\phi^{(n)}, \phi^{e(n)}]$
   - Compute actualized pickup probabilities $[P^{(n,k)}(z,t), P^{e(n,k)}(z,t)]$
   - Simulate resulting matches $m(z, t)$ and vacant cab distribution for $t + 1$
5: Approximate $\Gamma^{(n)}$ with the mean of the $M$ simulated actual ride probabilities $[P^{(n,k)}(z,t), P^{e(n,k)}(z,t)]$, compute objective series $F^{(n)} = \alpha \Gamma^{(n)} + (1 - \alpha) \Gamma^{(n-1)}$
6: Update convergence criterion $\|F^{(n)} - F^{(n-1)}\|$
7: Increment $n$, return to step 3 and continue to convergence
8: end while

1.3.4 Parameterization

To back out the data-generating equilibrium, I match the simulated number of street hail and e-hail matches, minimizing the norm of moment distance: $\min_{\tilde{P}, \tilde{P}^e} \|\hat{m}(\tilde{P}, \tilde{P}^e) - m\|$. The equation is exactly identified: the dimensions of the belief vector and the moment vector are both 5,760. Also, it is computationally infeasible to optimize over a high dimensional vector. Therefore, I make parameteric assumptions on the street hail and e-hail matching functions to reduce the dimension of parameter space. I assume a Cobb-Douglas street hail matching function and an independent-draw e-hail matching function over a dispatch network. See appendix C for more details on the functional forms.

Street Hail Matching Function

In aggregate, the street hail matching function with $C$ searching vacant cabs and $D$ waiting passengers is assumed to follow a Cobb-Douglas function: $m(C, D) = AC^a D^b$. 
I chose a Cobb Douglas functional form because simulations of moving searching cabs and waiting passengers in a rectangular lattice with geographical passenger arrival rate variation generate data is closely fitted by a Cobb-Douglas form matching function, as discussed in appendix C. From the perspective of each cab driver, the probability of getting a match would be \( \frac{AC^a D^b}{C_0^a} = AC^{a-1} D^b \). The event of a pick up is equivalent to the event of searching time less than the time he spends in zone \( z \). It is observed in the 2013 TLC panel trip record that distribution of searching times for drivers follows closely an exponential distribution. Therefore, the probability of matching with a passenger within \( x \) minutes is

\[
P(z, t|x; \lambda) = 1 - \exp(-\lambda(z, t)x) \tag{1.11}
\]

Assume that variation in passenger arrival rates across states is \( (z, t) \) absorbed by information of location and time, the probability a cab is matched within \( x(z, t) \) minutes given that it stays in zone \( z \) in period \( t \) can be expressed as:

\[
P(z, t) = 1 - \exp \left(-\lambda(z, t)x(z, t)\right) = C(z, t)^{\alpha z^{-1}} \beta z \gamma z, t \tag{1.12}
\]

**E-hail Matching Function**

Let \( \nu(z, t) \) be the probability that there is an e-hail cab within the matching distance mandated by TLC and \( \kappa \) be the average network strength defined as the number of cabs reached for each request.

\[
m^e(C^e, D^e) = \min\{C^e, D^e\}(1 - (1 - \nu(z, t))^\kappa) \tag{1.13}
\]

From the perspective of an e-hail driver the pickup probability is:
\[ P^e(z, t) = \min \left\{ 1, \frac{D^e(z, t)}{C^e(z, t)} \left( 1 - (1 - \nu(z, t))^\kappa \right) \right\} := \min \left\{ 1, \lambda^e(z, t) \left( 1 - (1 - \nu(z, t))^\kappa \right) \right\} \]

(1.14)

Derivation of the e-hail matching function is discussed in appendix C. In the reduced problem, drivers derive beliefs on pickup probabilities through street hail parameters \( \lambda^s \) and e-hail parameters \( \lambda^e \). Notice that \( \lambda^s, \lambda^e \) are functions of structural parameters \( \theta = (\{\alpha_z\}, \{\beta_z\}, \{\gamma\}_{z, t}, \kappa, \sigma, \sigma^e) \in \mathbb{R}^{44} \). In the simulated fixed point algorithm, instead of the belief domain \([0, 1]^{5760}\) I iterated on the parameter space in a much smaller domain \( \mathbb{R}^{44} \) as follows:

**Algorithm 2** Simulated Fixed Point Algorithm

1: Start with an initial structural parameter vector \( \theta \in \mathbb{R}^{44} \).
2: Evaluate beliefs \( \lambda(\theta) = (\lambda^s(\theta), \lambda^e(\theta)) \)
3: **while** non-convergence of beliefs **do**
4: Solve for Dynamic Programming decision rule \([\phi, \phi^e]\) given \( \lambda(\theta) \).
5: Simulate \( K \) paths of cab distribution \( C(z, t) \) and matches \( m(z, t) \) from the estimated shift starting time, zone and shift length distribution \( S_0 \).
6: Compute average actualized pickup probabilities \( \lambda' \) across simulated paths
7: Evaluate convergence criterion \( \|\lambda' - \lambda\| \)
8: Back to step 3 and continue to convergence
9: **end while**

Convergence of beliefs is satisfactory under the parameterization of matching functions. Conditional on the initial state \( S_0 \), I ran 20 iteration of simulations of \( K \) paths, where the belief of each iteration \( \lambda^{(n)} \) is computed as the mean of the \( K \) realized values of actualized \( \lambda^{(n,k)} \). At the limit, the mean absolute deviation between iterated beliefs is bounded above by 0.0005. With more simulations, the series converges at a faster rate, as shown in the figures below: the beliefs converges eventually within 10 periods, and higher number of simulations ran allows smaller convergence criteria for the same number of episodes.
1.3.5 Demand

I incorporate a discrete choice model on the demand side to estimate consumers’ response to price and waiting times and derive the consumer surplus functions. A
commuter $i$ who wants to travel from zone $z_1$ to $z_2$ at time $t$ has 3 choices: taxicabs, UberX ridesharing and MTA subway trains. The utility of choosing alternative $j$ is:

$$U_j^i = a_j X_j^i + b p_j + c E w_j + d T_j + \varepsilon_j^i$$  \hspace{1cm} (1.15)

Here I assume that the passenger has access to all three alternatives, and the unobserved utility $\varepsilon_j^i$ is generalized extreme value distributed. $p_j$ is the price of alternative $j$, $E t_j$ is the expected total travel time and $X_i$ consists of indicators of alternative $j$ and whether the trip originates or the trip originates and ends at different boroughs.

Specifically,

$$U_{\text{taxi}}^i = a_0 + a_1 I \{\text{crossboro}^i\} + b_1 P_{\text{taxi}}^i + c_1 E w_{\text{taxi}} + d_1 T_{\text{taxi}} + \varepsilon_{\text{taxi}}^i$$ \hspace{1cm} (1.16)

$$U_{\text{uber}}^i = a_0 + a_1 I \{\text{crossboro}^i\} + b_1 P_{\text{uber}}^i + c_1 E w_{\text{uber}} + d_1 T_{\text{uber}} + \varepsilon_{\text{uber}}^i$$ \hspace{1cm} (1.17)

$$U_{\text{mta}}^i = a_2 + c_2 E w_{\text{mta}} + d_2 T_{\text{mta}} + \varepsilon_{\text{mta}}^i$$ \hspace{1cm} (1.18)

There is no price component in the utility for riding metro because New York metro subway charges a flat fare of $2.75$ and I only consider metro riders who paid full fares. In terms of total travel time and pricing, passengers substitute taxicab and Uber closer to each other than to MTA. Specifically, the transportation choice problem is nested process: based on the trip characteristics (e.g. distance, weather) the passenger decides whether he will be choosing mass transit or ride-hailing services. Should he elect the latter, he then choose between taxicab and Uber. Let $k(j)$ denote the nest of alternative $j$,

$$U_j^i = a_{k(j)} X_j^i + IV_{k(j)} + \varepsilon_j^i$$ \hspace{1cm} (1.19)

Denote the variance of unobserved utility within nest $k(j)$ as $\xi_{k(j)}$, the inclusive value of nest $k(j)$ is:
\[ IV_{k(j)} = \max_{j \in k(j)} \left\{ U^i_j + \varepsilon^i_j \right\} = \xi_{k(j)} \log \left( \sum_{j \in k(j)} \exp \left( \frac{bp_j + cEt_j}{\xi_j} \right) \right) \] (1.20)

1.4 Estimation

1.4.1 Model Environment Parameters

\( x^t_z \) is the average time in minutes that a cab will stay at zone \( z \) during period \( t \) after dropping off a passenger at the same zone, regardless of the destination search zone the cab is going to next period. As suggested by Wang (2014), an approximation of \( x^t_z \) for zones that are small enough is the median searching time for trips right after a driver dropped off a passenger in the same zone that he picked up his next passenger. That is, \( \hat{x}^t_z = \text{median(searching time: dropped off at } (z,t), \text{ next trip at } z) \). The trip record data for year 2016 is an aggregate data, but trip record data with panel data structure was available for year 2013, which I used to find the inter-trip time and calibrate \( x^t_z \). For larger zones, I measure the physical dimensions in miles and imputed the average speed of cab movements at different time periods, and then found the implied required time to travel from the center of the zone to its boundaries. Another simplification is that \( \hat{x}^t_z \) is bounded above by 10. When \( x^t_z = 10 \), the cab stays in its originating zone for the entire period, and in the event of not getting a match, it will arrive at the destination zone next period. Table 1.7 shows the mean calibrated values of \( \hat{x}^t_z \) for each zone, and the standard deviation of the variation across \( t \).

The initial state for each driver consists of: his initial zone \( z_0 \), shift starting time \( t_0 \) and his shift length \( \tau_0 \). I assume that \( z_0 \) is the zone at which a driver first pick up a passenger in his shift, the probability mass distribution of \( z_0 \) is directly inferred from the 2013 dataset. And the average number of shifts started at each 15-minute
time period in 2016 is published by TLC in the 2016 taxicab and limousine indistry factbook.\textsuperscript{21} The joint distribution of \((t_0, \tau_0)\) is approximated by the empirical joint distribution observed in the 2013 dataset. \textsuperscript{22}

The 2013 TLC data set provides anonymous identifiers of medallions and driver hack licenses, and I identified shifts lengths as follows. Shift start and end times are identified when a mini-fleet cab have different drivers for consecutive trips. To identify shift endpoints for individual cab medallions, shift endpoints are defined if the time interval between two consecutive trips for the same driver are greater than or equal 4 hours.

\textsuperscript{22}While Buchholz (2016) and FLS (2017) consider labor supply endogenous and was modeled as an optimal stopping problem, I posit that the sampling from the empirical joint distribution will account for income shocks that determine labor decisions and therefore abstract away from this terminology.
Table 1.7: Estimated time a cab stays in zone where it dropped off a passenger ($\hat{x}$).

<table>
<thead>
<tr>
<th>Zone</th>
<th>Neighborhood</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bloomingdale, Manhattan Valley</td>
<td>2.93</td>
<td>0.46</td>
</tr>
<tr>
<td>2</td>
<td>Upper West Side N &amp; S</td>
<td>3.32</td>
<td>0.69</td>
</tr>
<tr>
<td>3</td>
<td>Lincoln Square E &amp; W</td>
<td>2.88</td>
<td>0.49</td>
</tr>
<tr>
<td>4</td>
<td>Clinton E &amp; W</td>
<td>3.04</td>
<td>0.68</td>
</tr>
<tr>
<td>5</td>
<td>E &amp; W Chelsea</td>
<td>3.38</td>
<td>1.22</td>
</tr>
<tr>
<td>6</td>
<td>Central Park</td>
<td>2.90</td>
<td>1.03</td>
</tr>
<tr>
<td>7</td>
<td>Midtown N &amp; S, Times Square</td>
<td>3.06</td>
<td>0.94</td>
</tr>
<tr>
<td>8</td>
<td>Garment &amp; Penn Station</td>
<td>2.43</td>
<td>0.78</td>
</tr>
<tr>
<td>9</td>
<td>Flatiron &amp; Union Square</td>
<td>2.55</td>
<td>0.68</td>
</tr>
<tr>
<td>10</td>
<td>E Harlem S</td>
<td>2.76</td>
<td>0.43</td>
</tr>
<tr>
<td>11</td>
<td>Upper East Side N, Yorkville S</td>
<td>3.36</td>
<td>0.78</td>
</tr>
<tr>
<td>12</td>
<td>Upper East Side S, Lenox Hill W</td>
<td>3.16</td>
<td>0.76</td>
</tr>
<tr>
<td>13</td>
<td>Midtown E, Sutton Place &amp; UN</td>
<td>3.06</td>
<td>0.78</td>
</tr>
<tr>
<td>14</td>
<td>Midtown S &amp; Murray Hill</td>
<td>2.71</td>
<td>0.75</td>
</tr>
<tr>
<td>15</td>
<td>Kips Bay &amp; Gramercy</td>
<td>2.91</td>
<td>0.61</td>
</tr>
<tr>
<td>16</td>
<td>Greenwich Village N, East Village</td>
<td>2.95</td>
<td>0.81</td>
</tr>
<tr>
<td>17</td>
<td>Alphabet City, Greenwich Village S &amp; West Village</td>
<td>3.11</td>
<td>1.25</td>
</tr>
<tr>
<td>18</td>
<td>Lower Manhattan</td>
<td>4.37</td>
<td>1.35</td>
</tr>
<tr>
<td>19</td>
<td>non-Manhattan zones</td>
<td>6.36</td>
<td>0.91</td>
</tr>
<tr>
<td>20</td>
<td>JFK Airport &amp; LaGuardia Airport</td>
<td>10.00</td>
<td>0</td>
</tr>
</tbody>
</table>

1.4.2 Payoff, Transition Kernel and Action Space for Drivers

The trip payoff and transition probability functions are inferred from data. As in the model each cab can take at most one trip for each period, the drop off point of each trip originating from $(z, t)$ will be $(z', \max\{t', t + 1\})$. The expected payoff and transition probabilities are inferred as the sample average for the observed trips with the same zone-period pair end points. The payoff obtained by a driver for a trip is measured as the total amount (fare plus tip). In the model, I restrict that trips must be at least 1 period long (10 minutes) so states are not recurrent. To account for trips that pick up and drop off took place within the same time period $t$, I assign the
transition probability of the trip \((z, t) \rightarrow (z', t)\) to the trip \((z, t) \rightarrow (z', t + 1)\), and I aggregated the amount from that trip to the driver’s next trip.

The action space is assumed to be time-invariant. For each zone, there are at most 7 possible zones a driver can go to, including mostly their adjacent zones for zones in Manhattan. For zones that are near bridges and tunnels that lead cabs out of Manhattan, the option of going to other boroughs and airports will be available. Likewise, from the non-Manhattan zones (other boroughs and airport), the options of going back to these zones will be available. The time travel costs for adjacent zone pairs are zero, while those for non-adjacent zones are approximated with the trip time with the corresponding trip endpoints at the same time period. I compared my action space setting to the empirical searching probabilities observed with more than 5% frequencies. Almost all the zone pairs are consistent. This also shows that 10-minute time intervals would be a reasonable partition to capture drivers’ movements without having a very high dimensional action space.

1.4.3 E-hail Market Parameters

The number of vacant green cabs are published in the fact book. The fact book does not include the percentage of drivers that are equipped with e-hail applications for both yellow and green cabs. The number of e-hail drivers were included in the TLC e-hail pilot program quarterly reports.\(^{23}\) I extrapolate the number of e-hail drivers and approximately 25% of yellow cab drivers are equipped with e-hail applications. The percentage of e-hail green cab drivers is a parameter to be estimated. The arrival and cancellation rates of e-hail requests are directly inferred from data as the average number of request arrivals and cancels for all \((z, t)\) pairs. The trip completion probability of the e-hail requests and e-hail passengers is estimated by using the TLC e-hail

guideline that the distance between the request and cab needs to be within 0.5 miles in
downtown Manhattan and 1.5 miles elsewhere, and the fact that a zone in the model
is in fact a combination of smaller subzones (TLC zone partition). The probability
that an e-hail cab in zone \( z \) is near the location of the request is approximately the
probability that the cab is in the same subzone \( z_s \) that the request originates from,
which is \( \sum_{z_s \in z} p_C(z_s|z)p_D(z_s|z) \). The probability \( p_D(z_s|z) \) is measured as the relative
frequency, and I approximate \( p_C(z_s|z) \) as the relative pickup frequency from subzone
\( z_s \) for street hail rides in \( z \), which encompasses the search behaviors within a zone.

1.4.4 Matching Function Parameters

The estimation of structural parameter vector follows the structure of nested fixed
point algorithm in Rust (1987) with a computation of equilibrium given the value
of structural parameter in the inner loop. I use simulated fixed point algorithm to
solve for the equilibrium decision rule can be when the iteration converges. Drivers
have a common initial belief on the probability of getting a ride at each state through
the \( \lambda, \lambda^e \). Based on their belief on the drivers maximize their dynamic payoff for
the shift with respect to the Bellman equations for street-hail and e-hail drivers. The
optimal policy of each driver \( \phi \) and \( \phi^e \) induces the spatial distribution and aggregate
law of motion of cabs, that altogether generate the actualized \( \lambda, \lambda^e \) and the resulting
number of matches, which are then matched with the sample moments at convergence.
Denote parameter vector \( \theta = (\{\alpha\}_z, \{\beta\}_z^t, \{\gamma\}_z^t, \kappa, \sigma, \sigma^e) \) the structural parameters in
the matching functions, the below algorithm describes the outer loop of the nested
fixed point estimation algorithm.
Algorithm 3 Simulated method of moments (Outer loop)

1: Sample an initial state matrix $S_0$ and compute actual matches $E[m_z^t]$ for each $(z,t)$ pair under initial parameter guess $\theta_0$
2: Evaluate moment distance of using simulated and actual matches
3: while moment distance $> \varepsilon$ do
4: Solve for equilibrium with the fixed point algorithm with parameter $\theta$
5: Simulate the number of matches produced $m$
6: Compute weighted moment distance and evaluate $\varepsilon$-convergence criterion
7: Update $\theta$
8: Back to step 4 and continue to convergence
9: end while

1.5 Results

1.5.1 Matching Parameters

The structural parameters $\theta$ is estimated with the method of simulated moments, where the moments matched are the number of matches for each (zone, period) pair, $m_z^t$. In each iteration 15 simulations were run to obtain the simulated number of street hail and e-hail matches produced at each (zone, time) pair. There are 5,760 moment restrictions for 44 parameters. To check if the model fits well on unmatched moments such as shift earnings, searching times and utilization rates of drivers, the estimation is also run with the 2013 panel dataset, as the corresponding data are unobserved for the 2016 data. There are 44 parameters that are estimated: 40 street hail matching parameters, variances of the extreme value alternative specific idiosyncratic shocks for street hail and e-hail drivers, e-hail app network factor and percentage of e-hail green cabs. The table for estimates of the parameters is included in Appendix A.

1.5.2 Fit of Model

The figures below show the fit of the model in terms of number of street hail and e-hail matches produced in each zone at different time periods in a typical weekday. The red
dotted lines correspond to the sample average of matches occurred in the same \((z,t)\) pair. With the estimates obtained from the simulated generalized method of moments, the model manages to capture the trends in most of the zones, succeeding in matching the peaks and troughs with the variation in time.\textsuperscript{24} The average absolute deviation of simulated number of street hail and e-hail matches from the sample moments are 25.7 and 0.79.

Parameters were estimated with the 2013 panel data, the model also fits the actual unmatched moments very well. In particular, I compared simulated and actual shift earnings, number of rides per shift and utilization rates. As shown in the table below, the model fits the unmatched moments fairly well. The histograms for distribution of earnings and utilization rates are shown below, the model predicts a slightly lower utilization rates for drivers, even though it is possible that the rather high utilization rates were due to measurement error or misreporting, it is also likely that there was a decrease in taxicab demand in 2016 due to the growing popularity of peer-to-peer ridesharing services such as Uber.

\textbf{Table 1.8:} Shift descriptive statistics on model fit (2013).

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Simulation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total pickups</td>
<td>395,720</td>
<td>385,880</td>
<td>(5,624)</td>
</tr>
<tr>
<td>Per hour shift earnings</td>
<td>41.03</td>
<td>38.319</td>
<td>(4.73)</td>
</tr>
<tr>
<td>Average trip per shift</td>
<td>20.92</td>
<td>18.25</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Searching time per shift (hours)</td>
<td>3.12</td>
<td>4 (0.062)</td>
<td></td>
</tr>
<tr>
<td>Utilization rate of shifts</td>
<td>60%</td>
<td>52.66%</td>
<td>(0.7%)</td>
</tr>
</tbody>
</table>

\textsuperscript{24}The model does not perform very well for the airport zone since the afternoon, and it does not succeed in replicating the smooth pattern during the witching hour. An investigation into the data set suggests that the pick up smoothing effects were generated by behavior of individual medallion drivers. A possible explanation that the model does not predict well in the airport is that these drivers are more likely to take very long shifts and are not considered in the model.
Figure 1.12: Simulated and true distributions of shift earnings.

Figure 1.13: Simulated and true distributions of utilization rates.
Figure 1.14: Simulated and true street hail matches.
Figure 1.15: Simulated and true e-hail matches.
Fig. 1.16 shows the value function $V$ across all zones in the model at the beginning of a 12-hour shift. Notice that even though the value function peaks at 12pm, implying that the value of a single shift is highest when it starts at 12pm, the optimal shift starting times would be around 6am and 6pm. If the medallion owner considers the aggregate value of the two shifts, the value is maximized when the shift changing times are set around 6am and 6pm. Fig. 1.17 shows the aggregated value of search for the two shifts in a 24-hour cycle. Unlike Buchholz (2018) who focuses on day shift drivers, the order of value functions across locations varies with time.\textsuperscript{25}

\textsuperscript{25}For example, the value function of Greenwich village is higher than Yorkville in the morning but lower in the afternoon.
Figure 1.16: Value function across all locations.
1.5.3 Demand Parameters

The demand parameters were estimated using a set of imputed datasets. A section dedicated to discuss in detail imputation procedures are included in the Appendix E. In the imputation process, four major assumptions are made:

1. The surveyed origin-destination pairs observed for weekday trips in the 2008 MTA New York City Travel Survey are representative on the MTA riders in the estimation dataset.

2. The origin-destination-hour distribution of 2016 Uber riders are the same as the same distribution for 2013 taxicab riders.

3. The trip distance duration is for taxi and Uber trips are drawn from the identical normal distribution, controlling for origin, destination and hour of pick up.
4. The waiting times for taxicab riders are exponentially distributed with parameter (mean waiting time) that is implied by competitive equilibrium and heat mapping motions within zones. (see Appendix F)

I estimated demand parameters over 100 samples each containing 100,000 passengers in the data and the table below shows the mean of the of parameter estimates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Point estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>base utility for ride</td>
<td>10.365</td>
<td>1.991</td>
</tr>
<tr>
<td>$a_1$</td>
<td>base utility for cross-boro ride</td>
<td>−4.364</td>
<td>0.551</td>
</tr>
<tr>
<td>$a_2$</td>
<td>base utility for metro</td>
<td>−85.34</td>
<td>4.4</td>
</tr>
<tr>
<td>$b_1$</td>
<td>coefficient of price for taxi and UberX</td>
<td>−14.520</td>
<td>2.467</td>
</tr>
<tr>
<td>$c_1$</td>
<td>coefficient of waiting time for taxi and UberX</td>
<td>−7.783</td>
<td>3.441</td>
</tr>
<tr>
<td>$c_2$</td>
<td>coefficient of travel time for metro</td>
<td>−1.188</td>
<td>0.288</td>
</tr>
<tr>
<td>$d_1$</td>
<td>coefficient of travel time for taxi and UberX</td>
<td>−2.873</td>
<td>0.120</td>
</tr>
<tr>
<td>$d_2$</td>
<td>coefficient of travel time for metro</td>
<td>−1.147</td>
<td>0.116</td>
</tr>
<tr>
<td>$\xi$</td>
<td>dissimilarity parameter for the ride hailing nest</td>
<td>0.549</td>
<td>0.362</td>
</tr>
</tbody>
</table>

Referring to the structural utility equations (1.3.5), the fixed utility of riding taxi and Uber relative to metro is positive, and it decreases for cross borough trips. Also, price sensitivity for metro is higher than that of ride hailing services, while travel time sensitivity for ride hailing services is higher than metro. The relatively small value of $\xi$ suggests a high substitutability between taxi and Uber.

In particular, that under the available options, the average willingness to pay for each reduced minute of travel time in terms of taxi fare is the ratio of waiting time coefficient to the price coefficient for the ride alternative nest, which is $0.54. This implies that the current booking surcharge of $2 under the UberTAXI platform may not attract passengers to switch from hailing cabs on the street to requesting over the app unless the waiting time reduced by 3.7 minutes. I collected the waiting time data of UberTAXI and compared with the approximated waiting times of taxi hailing from the streets.
Table 1.10: UberTAXI waiting time data.

<table>
<thead>
<tr>
<th></th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.59</td>
<td>4.47</td>
<td>4.66</td>
<td>4.71</td>
<td>4.28</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>2.27</td>
<td>2.29</td>
<td>2.3</td>
<td>2.34</td>
<td>2.04</td>
</tr>
<tr>
<td>Null entries</td>
<td>104,426</td>
<td>309,367</td>
<td>344,024</td>
<td>77,189</td>
<td>14,262</td>
</tr>
<tr>
<td>Total</td>
<td>181,958</td>
<td>527,193</td>
<td>584,223</td>
<td>128,343</td>
<td>22,698</td>
</tr>
</tbody>
</table>

Figure 1.18: Distribution of UberTAXI waiting time.

When there are UberTAXI partner cabs nearby the passenger, UberTAXI estimates that most requests can be matched within 5 minutes, as shown in the histogram. Note that the distribution of UberTAXI waiting time is a truncated distribution and observation is conditional on positive response, or the event that at least a partner cab is sufficiently close to where the request was sent. However, according to our estimates in Manhattan (yellow line), the waiting time for UberTAXI (black dotted line) is higher than street hail waiting time except during the morning and evening peaks in certain zones. This suggests reasons that the low number of requests observed is due to the fact that there is not much waiting time improvement for taxi services. Even when there is, the magnitude of improvement is not big enough to outweigh the platform booking surcharge.
Figure 1.19: Street hail (solid) and UberTAXI (dotted) waiting time in Manhattan.

1.6 Counterfactual Experiments

Using the supply and demand side estimates from my model, I consider two sets of counterfactual experiments. In the first set of experiments, I explore the efficiency gain of using e-hail matching. First, I assume that the demand pattern is constant across the locations and I substitute a fixed percentage of street hail demand exogenously, without considering incentives from passengers to demonstrate the gain in efficiency by using centralized platforms of matching. In the second experiment, I conduct an experiment of a universal e-hail dispatcher with no booking fees, relaxing the exogenous demand assumption in the first set. After these changes in demand system, I allow passengers and drivers to re-optimize under changes in model environment and study the equilibrium outcomes.\footnote{A limitation of these experiments is a lack of general equilibrium structure; neither the supply and demand side consider the impact of their actions on the other side. Chapter 2}
effects of using a flexible leasing schedule. An experiment of carpooling applications on e-hail is included in Appendix B.

1.6.1 Efficiency of E-hail Matching

Substituting Street Hail Demand with E-hail

This experiment compares e-hail matching with the estimated street hail function, holding the demand pattern. Under each of the simulations below, the street hail demand parameters of every zone are decreased by a given percentage, and the implied number of matches substituted are added to the e-hail requests. In each of these cases and I assume the percentages for yellow and green e-hail drivers grow by the same percentages.\(^{27}\) I also kept constant the cancellation rates and e-hail meeting rates \(\nu\).

Under these conditions, I allow drivers to re-optimize under the experimental model environment. The table below presents the simulated equilibrium outcome in each of the cases.

As seen Table 1.11, when passengers substitute to use e-hail apps, the earnings and utilization for e-hail drivers increases, at the expense of the non-e-hail drivers. Also, now that Manhattan areas also have high enough e-hail request arrival rates, e-hail yellow cab drivers are incentivized to search in Manhattan, decreasing the overall rides outside Manhattan. Also note that under the current e-hail app network factor there is a capacity and limit for earning and utilization improvements.

On the other hand, the e-hail premium decreases as there are more drivers using e-hail programs. The below table shows the differences in shift earnings for e-hail and

\(^{27}\)With a growing network of passengers, I also allowed the e-hail app network strengths \(\kappa\) is set to 10. The estimated \(\hat{\kappa}\), number of drivers reached by each request is 5.6. I allow the network strength to grow to 10 in each of these simulations.
**Table 1.11:** Equilibrium outcomes across percentage of street hail demand substituted.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>% of substituted rides</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30%</td>
</tr>
<tr>
<td>Shift earnings (non-e hail)</td>
<td>$232.02</td>
</tr>
<tr>
<td>Shift earnings (e hail)</td>
<td>$243.47</td>
</tr>
<tr>
<td>Earnings per hour (non-e hail)</td>
<td>$26.32</td>
</tr>
<tr>
<td>Earnings per hour (e hail)</td>
<td>$27.58</td>
</tr>
<tr>
<td>Utilization rate (non-e hail)</td>
<td>48.93%</td>
</tr>
<tr>
<td>Utilization rate (e hail)</td>
<td>49.13%</td>
</tr>
<tr>
<td>Total street hail rides</td>
<td>317,800</td>
</tr>
<tr>
<td>Total e-hail rides</td>
<td>2,417</td>
</tr>
<tr>
<td>Total rides in non-yellow zones</td>
<td>20,195</td>
</tr>
</tbody>
</table>

non e-hail drivers, assuming a 20% increase in e-hail demand with respect to the data. As the percentage of drivers using e-hail programs increases, market tightness increase and e-hail drivers face more competition. Eventually, as there are more e-hail drivers, the premium decreases, and shift earnings approach those of non-e-hail drivers.

**Table 1.12:** Earnings for e-hail vs non e-hail drivers.

<table>
<thead>
<tr>
<th>Percentage of e-hail drivers</th>
<th>10%</th>
<th>30%</th>
<th>50%</th>
<th>70%</th>
<th>90%</th>
<th>non-e hail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shift earnings</td>
<td>$364.86</td>
<td>$297.55</td>
<td>$271.95</td>
<td>$260.53</td>
<td>$253.21</td>
<td>$225.15</td>
</tr>
<tr>
<td>Per hour earnings</td>
<td>$41.48</td>
<td>$33.75</td>
<td>$30.87</td>
<td>$29.53</td>
<td>$28.73</td>
<td>$25.58</td>
</tr>
<tr>
<td>Utilization rate</td>
<td>74.14%</td>
<td>60.69%</td>
<td>55.73%</td>
<td>53.60%</td>
<td>52.00%</td>
<td>48.50%</td>
</tr>
<tr>
<td>Total e-hail pickups</td>
<td>33,664</td>
<td>52,175</td>
<td>53,011</td>
<td>53,378</td>
<td>53,627</td>
<td>—</td>
</tr>
<tr>
<td>Portion of e-hail trips for each</td>
<td>67.84%</td>
<td>42.54%</td>
<td>29.22%</td>
<td>22.09%</td>
<td>15.93%</td>
<td>—</td>
</tr>
</tbody>
</table>

**A Universal, City-Sponsored E-hail Platform**

As shown in the experiment above, e-hail matching is more efficient than street hail matching. Comparing with traditional street hail matching function, drivers are able
to utilize their shift time better and match with a higher number of passengers, given the estimated exogenous net demand arrival process. Our next experiment considers the effect of making the e-hail platform a universal dispatcher through a city-sponsored e-hail application, taking into account response from the demand side. Specifically, I use the discrete transportation choice model for passengers to derive the effect on consumer surplus when the city sponsors all taxicab drivers to use a universal e-hail dispatcher app. By doing so the network strength of e-hail cabs is increased significantly, and I waive the $2 booking fee.

For passengers, the main improvement from using e-hail is reduced waiting times. To model the relationship between waiting time and market thickness across all locations, I use a grid world simulation based on the waiting time simulation introduced in Lizzeri et al (2017). Details of waiting time simulations inside and outside Manhattan is included in section F of the Appendix. As shown in the two figures below that compares the waiting time of taxicabs with Uber in Manhattan before and after the implementation of the universal e-hail dispatcher, there is a significant improvement in waiting times under e-hail and gives taxicab an advantage against Uber.
Figure 1.20: Simulated waiting times for street hail taxicabs (solid) vs. Uber (dotted) in Manhattan.
Figure 1.21: Simulated waiting times for taxicabs (solid) vs. Uber (dotted) in Manhattan under universal dispatcher.

I calculate the gains in consumer surplus for passengers in terms the expected willingness to pay: \( \sum_i \left( P^i(taxi) \times WTP(\Delta Et^i) \right) \). The estimated gains in consumer surplus for each commuter is $0.15, which amounts to $33,698 per day or $25.32 million per year. Without the booking fee, the improved e-hail network strength reduces total travel time by 2.17 minutes, which is expected to induce an additional 21,060 taxicab trips substituted from ride sharing and metro per day. On the supply side, when drivers re-optimize given the change in demand, average daily shift income increases by $13 and their utilization rate increases by 10%.
1.6.2 Flexible Leasing Schedules

Surplus of Flexibility under the Uber Shift System

I employ the terminology in Chen et al. (2015) to model drivers’ decisions under the flexible leasing program. Suppose drivers’ reservation wage at hour $h$ in a day $d$ is given by:

$$w_t^* = \mu_{h,d} + \xi_{h,d} \tag{1.21}$$

Where $\mu$ is the deterministic component, observable by the econometrician while $\xi$ is only observed by the driver. Furthermore, the unobserved component can be decomposed as the sum of a day shock and an hour shock:

$$\xi_{h,d} = \nu_h + \nu_d \tag{1.22}$$
$$\nu_d \sim N(0, \sigma_d^2) \tag{1.23}$$
$$\nu_h \sim N(0, \sigma_h^2) \tag{1.24}$$

Under the 12-hour requirement, the driver only conditions on the day shock. If $V(h,d)$ is the net-of-expense value of driving for hour $h$, he chooses to take the shift if $\sum_h (V(h) - \mu(h)) > \sum_h \xi_h$. Before the driver makes the decision, he observes the realized value of the daily shock, $\nu_d$ and he has a rational expectation on the hourly shock $\nu_h$. Under the flexible leasing rule, the expected surplus is

$$S_s = \Phi \left( \frac{\sum_h (V(h) - \mu(h))}{\sqrt{12} \sigma_h} \right) \left( \sum_h (V(h) - \mu(h)) + \sqrt{12} \sigma_h \frac{\phi \left( \frac{\sum_h (V(h) - \mu(h))}{\sqrt{12} \sigma_h} \right)}{\Phi \left( \frac{\sum_h (V(h) - \mu(h))}{\sqrt{12} \sigma_h} \right)} \right) \tag{1.25}$$

---

28 Chen et al. (2015) assumes that drivers cannot foresee shocks, here I abstract from this
If the expression on the right hand side is negative, the driver elects to not commit to the 12-hour shift and derives zero surplus. Under the ideal flexible rule in which drivers have full control on shift schedules, he chooses to work in hour \( h \) if \( V(h, d) > w^*_h,d \), the expected surplus for the hour is and the same 12 hours are:

\[
S_h = \Phi\left(\frac{V(h) - \mu(h)}{\sigma_h}\right) \left(V(h) - \mu(h) + \sigma_h \frac{\phi\left(\frac{V(h) - \mu(h)}{\sigma_h}\right)}{\Phi\left(\frac{V(h) - \mu(h)}{\sigma_h}\right)}\right)
\]

\[
S_{fs} = \sum_h \max\{S_h, 0\}
\]

It can be shown analytically that \( S_{fs} \geq S_s \). I simulate a path of reservation wage and study the gain in surplus when a cab driver operates under the Uber shift system, where he makes hourly entry and exit decisions. The reservation wage is modeled as the sum of mean reservation wage and the hour shock. The idiocyncratic mean reservation wage is not identified, as panel data for cab drivers is not available. Instead I took the reservation wage as $17.22, the suggested hourly net-of-expense minimum wage paid to all for-hire vehicle drivers in New York (Parrott and Reich (2018)) - if the driver owns a car he can join the ride sharing network when he exits the market. On the other hand, I assumed that standard deviation of the hourly shock follows a log normal distribution: \( \sigma_h \sim \exp(N(1, 0.5)) \). The lognormal parameters for are taken from Chen et al. (2015) and these parameters fit the labor supply decisions of Uber drivers. The shift leasing price for each hour is taken as the leasing cap set by TLC.\(^{29}\)

Figure 1.22 shows the median, 5th and 95th percentiles of simulated total expected surplus in a 12-hour shift started at each hour if cab drivers operate under an Uber style shift system. When the cab driver is able to enter and exit the market anytime

during his shift, he gains an average extra surplus from $3.7 to $8.1, or $25 to $50 per week. The variation in surplus in this simulation is induced by randomness of reservation wages, and drivers assume that they earn the average hourly fare for taxi when they enter the market and the suggested minimum wage for a New York for-hire vehicle driver.

Figure 1.22: 5th and 95th percentiles of simulated total expected surplus in a 12-hour shift (by shift starting hour).

Aggregate Effects

It is noted that the Uber labor supply model may not be feasible as this stage because full freedom on entry and exit requires that drivers own their vehicles. In reality, drivers do not necessarily own the taxicab they drive. In the counterfactual simulation drivers are assumed to decide how many hours to work for one continuous shift and when to start their shifts given the working hours. Drivers’ decisions on driving hours depend on the disutility function for work. Since the data used does not provide
driver level variables, disutility function parameters cannot be identified. The counterfactual simulation assumes the distribution of working hours reveal information about drivers’ preference for driving hours and it does not differ from the data. Condition on the working hour drawn a driver chooses the starting time for his shift considering potential reservation wage shocks. The graphs below display the number of actives cabs and number of matches at each time period. Under the flexible leasing schedule, the reduction in the number of cabs on the road during the evening rush hour is avoided. The average earnings per shift increases from $271.2 to $278.4, resulting from 3,710 additional completed rides.

**Figure 1.23:** Active cabs (solid: flex leasing, dotted: current).
1.7 Conclusion

This paper presented a dynamic spatial matching game model for yellow medallion cab drivers. In equilibrium, pure street hail and e-hail drivers make optimal spatial search decisions and their strategies are consistent with the rational expectations. To account for demand side response, I constructed a data set from taxicab trip records, Uber’s pricing and ETA data and metro subway ridership records to estimate passengers’ price and waiting time elasticities.

Counterfactual experiments show that using e-hail technologies generates gains in drivers’ income and utilization when e-hail demand substitutes street hail demand. Also, waiving the $2 booking surcharge required by TLC and improving the e-hail network strength generate consumer surplus of $25.32 million per year. Finally, allowing flexible shift schedule allows drivers to respond to short term reservation shocks, providing weekly surplus around $25 with respect to the for-hire vehicle minimum wage. When implemented, the drop in the number of active cabs during the evening peak is avoided, creating 3,710 additional matches and also improved drivers’ earnings.

Figure 1.24: Completed trips (solid: flex leasing, dotted: current).
Under the current regulations set by TLC taxicabs have been losing market shares to peer-to-peer ridesharing services such as Uber. Competition across alternatives are determined by prices and waiting times. E-hail technology reduces waiting times, but this improvement depends on having a sufficiently large number of e-hail drivers. A counterfactual experiment in FLS which they predict market segmentation on the e-hail dispatching platform causes the income of e-hail dispatch drivers to decrease. My result is in the same spirit, note that e-hail drivers can still pick up street hail passengers, market segmentation over dispatchers causes average network strength to be weak, inducing little gains (but not losses) in drivers’ earnings and matches produced. Along with improved network strength, waiving the required booking surcharge can improve the competitive edge of e-hail taxi on the price dimension, generating substantial consumer surplus gains. This indicates a policy of using a centralized e-hail dispatch platform could revolutionize the taxicab industry: a city-sponsored, universal dispatch platform may be optimal rather than the current oligopolistic e-hail taxi dispatcher market regime.

TLC is experimenting with other pilot programs to improve the taxicab market in New York City, including the flexible fare pilot which allows e-hail dispatchers to charge a fare different from the standard fare meter rate. Chapter 2 extends the structural model of taxicab searching and matching presented in this chapter to study the welfare effects of using two candidates of dynamic pricing strategies, including spatial and surge pricing.
Chapter 2

Dynamic Pricing and Cross-Platform Competition in the NYC For-Hire Vehicle Industry

2.1 Introduction

Mobility-on-demand systems have improved the for-hire vehicle market, mainly in the form of centralized peer-to-peer matching platforms. These platforms, operated by for-hire vehicle dispatchers, are available to network large numbers of users through mobile-phone applications that match riders with passengers, providing more efficient matching mechanisms by reducing drivers’ searching times and passengers’ waiting times. Also essential in the for-hire vehicle industry is dynamic pricing. Price regulation is heterogeneous. Peer-to-peer ride-sharing platforms including Uber and Lyft face little restrictions on setting the fare for each individual trip, and they are free to set fare rates conditional on pickup time and location, trip distance and duration, and individual discount offers. On the other hand, e-hail taxi dispatchers, who partner with medallion taxicab drivers, were restricted to charging the regular taxi meter fare until the August 2018 launch of a pilot program for flexible fares, which has allowed for more flexibility for dispatchers to set fare rates.¹

Peer-to-peer ride-sharing platforms are well known for their use of dynamic pricing schedules in response to changes in supply and demand for rides. For example, Uber and Lyft adopt dynamic pricing schemes known as surge pricing and prime time for

¹As of April 2019, there are only 3 approved participants: Waave, Wapanda, and Myle.
their ridesharing services. As of the time frame in which the data set used in this paper was collected, this scheme is implemented as Uberâ€™s system computes a surge multiplier in \([1, \infty)\), which is rounded to 1 decimal place based on real-time market parameters and then multiplied to the standard fare rate on the basis of distance and trip duration.\(^2\) Surge multipliers are known as prime-time factors in Lyft’s case. According Uber’s and Lyft’s webpages, the adoption of surge pricing is both to ensure that passengers with highest willingness-to-pay get matched with drivers\(^3\) and to boost drivers’ earnings\(^4\). In practice, drivers can observe surge multipliers in real time. On occasions including major holidays, promotions, or events, surge-pricing schedules at given times and locations are guaranteed and revealed to drivers \textit{ex ante} to their labor decisions (e.g. the Guaranteed Prime Time program for Lyft drivers), so that they can plan their shift times in advance.

Several seminal works studying the effects of dynamic pricing in the for-hire industry are based on Uber surge-pricing data provided by or collected from Uber. With access to Uber’s internal driver data, Chen and Sheldon (2015) and Hall et al. (2015) have studied the effects and mechanism of surge pricing on both supply and demand sides. Chen and Sheldon (2015) show that surge pricing induces prolonged driving hours for Uber partner drivers; Hall et al. (2015) show that Uber’s surge-pricing algorithm both filters excess demand and encourages supply in surging areas so that the platform allocates rides to riders with highest willingness-to-pay, while maintaining sufficiently low waiting times for them. Cramer and Krueger (2016) have offered a qualitative argument on how surge pricing leads to overall welfare gains as

\(^2\)Instead of reporting the surge multiplier to the passenger, Uber currently quotes an upfront pricing when a trip request is made. The upfront price implicitly considers the surge multiplier and the expected travel time. See https://www.uber.com/ride/how-uber-works/upfront-pricing/.

\(^3\)https://www.uber.com/drive/partner-app/how-surge-works/.

it matches supply and demand more closely throughout the day. Castillo, Knoeefle and Weyl (2017) have studied trip data provided by Uber and found evidence of the wild goose chase phenomenon in Manhattan: in their study, low prices decreased the number of available cabs, which on the one hand increased waiting time for passengers and on the other hand decreased utilization rates for drivers as the distance between drivers and passengers increased. The study also showed quantitatively that surge pricing improved total social welfare.

Taxi shift hours are more rigid than hours for Uber drivers (Chen 2017), but real-time dynamic pricing can still be effective. Not only does real-time dynamic pricing incentivize prolonged driving hours and drivers’ entry, but Chen, Mislove, and Wilson (2015) have also found evidence that it spacially redistributes existing supply. In particular, they collected real-time location data of Uber drivers through pingClient and found that, when an area had a price surge relative to its neighboring areas, trajectories of vacant Uber vehicles in neighboring areas suggested that these drivers left their originating locations to search for rides in the surging area. In line with the findings of Chen, Mislove, and Wilson (2015), I have found empirical evidence consistent with what the reduced-form model on Uber waiting time predicts for spatial redistribution of existing Uber vehicles during a price surge. When surge multipliers in an area are controlled for, the average lagged surge multipliers in a neighbor area have a significant positive effect on the area’s waiting time.

In this paper, I extended the structural model of the taxicab market in chapter 1 to study the welfare effects of implementing dynamic pricing on e-hail taxi dispatch platforms, taxi drivers, and riders under various pricing schedules and market setups. In this model, each cab driver belongs to exactly one e-hail network, while passengers practice multihoming and choose which e-hail platform to hail. This is an extension of chapter 1’s model, since in supply-demand equilibrium drivers and passengers endo-
genize the responses of their collective decisions on the other side. Managers of the dispatchers consider the supply-demand equilibrium outcome under current prices set by them and their competitors, and thus platforms optimize their welfare functions. Restrictions on shift schedules give taxicab drivers less flexible working hours than partner drivers have on ride-sharing networks. Therefore, I treated entry and shift hours of taxi drivers as exogenous, so pricing schedules’ effects on market tightness across locations channel through the spatial-redistribution effect.

As noted by Economides (1995) and Tirole and Rochet (2003), e-hail platforms exhibit network externalities. Supply and demand sides each contribute to the surpluses of the other side. Higher demand at a location increases pickup probabilities and therefore drivers’ value of search in that location. Also, large network size of drivers implies smaller distances between drivers and passengers and therefore lower waiting and searching times (Castillo et al. 2017). On the other hand, profit-maximizing behavior of platforms may lead to non-trivial welfare changes for riders: large demands may allow higher markups for platforms with large networks. In this model, I studied duopoly setups across various degrees of market concentration to study how market power and competition affect prices and consumer surpluses.

To my knowledge, this is among the first papers on welfare effects of dynamic pricing in the for-hire vehicle industry which takes into account competition between dispatchers. Lee (2017) includes a simulation-based study on the competition of ride-sharing platforms under a hypothetical Salop (1979) circle setup in which distribution of drivers is exogenous and passengers respond only on prices. I considered my model in an empirical context, namely New York City. As noted by Armstrong (2006), competition in two-sided platforms takes place across and within networks. Therefore, it is crucial to consider both passengers’ endogenous responses to platforms’ pricing rules and drivers’ strategies in modeling rail-hailing platform competition. In this model,
decentralized drivers respond to their own beliefs of supply and demand, and passengers respond to prices and expected waiting times across platforms. In the competitive equilibrium defined, the model considers intra- and inter-platform competition explicitly through the prices, waiting times, and probability of matching. This general equilibrium approach allows for quantitative estimations of welfare effects for drivers, passengers, and the platform which are robust to agents’ optimizing behaviors.

To study welfare effects quantitatively under dynamic pricing schedules, I calibrated the structural model with a 3-location setup featuring Manhattan midtown center, east, and west, and I estimated waiting time as a function of tightness in the area using an approach similar to the grid map simulation in chapter 1. Given a price vector, one can estimate the supply-demand equilibrium using a nested variation of the iterative method presented in chapter 1. Successively approximating beliefs, using the contraction mapping property of the logit function, achieves convergence. In contrast to the estimation in chapter 1, in which net demand is exogenous, multiple supply-demand equilibria exist when the e-hail market is a duopoly. Throughout this chapter, I conditioned my analysis on the equilibrium selection rule that would induce equal market shares when the numbers of drivers on each platform were equal.

In the simulation experiment section, I considered two batches of experiments. In the first batch, I assumed that the e-hail taxicab market is run by a monopolist, and I considered welfare changes caused by dynamic pricing schedules when the monopolist is maximizing profit and when the monopolist is maximizing total social welfare. In the second batch, I consider a duopolistic e-hail taxicab market and study the optimal pricing rules for each of the two platforms across various degrees of market concentration. The two pricing schedules I considered are a predetermined region-specific multiplier and a surge-pricing algorithm that responds linearly to real-time demand-to-supply ratio.
The benchmark results demonstrate that the e-hail taxicab market exhibits network externalities under flat-rate pricing: when the e-hail taxi market is operated by a monopolist, both surplus for the consumer and earning for the platform are higher than under a duopoly. In the dynamic-pricing experiment under monopoly, I found that all dynamic pricing schemes improve social surplus. In particular, real-time surge pricing generates a mutual gain for the platform and riders; in contrast, regional pricing gives the platform a gain in earnings, but the platform maximizes its profits at the expense of consumers.

Dynamic-pricing experiments under duopoly shed some light on the nature of competition in the for-hire vehicle industry. Dispatchers of different shares of drivers compete against each other not only on fare rates but also on waiting times and request fulfillment rates, both of which are determined by the market thickness in the location considered. Simulation results show that, under a demand-side equilibrium response, surge pricing is rarely triggered, so competition outcome is determined mainly by the network size advantage. When surge pricing is adopted, a large platform utilizes its large network size and gains market dominance across all locations. On the other hand, under regional pricing, an inferior firm is able to set prices to support a zonal equilibrium (when the number of cabs is feasible). In any case, surge pricing generates a higher total social surplus.

The rest of the paper proceeds as follows. Section 2 describes the data used in this study and empirical findings that motivate the model. Section 3 provides details on the model. Section 4 defines the competitive equilibrium. Section 5 describes the estimation methods. Section 6 presents and discusses the counterfactual results. Section 7 concludes.
2.2 Empirical Findings in Data

To study the effects of surge pricing, I collected and analyzed real-time data on surge pricing and waiting time from the official Uber application programming interface (API). As described in chapter 1, I scripted an automated collection of data across 263 locations in New York City. This map partition is defined by the Taxicab and Limousine Commission (TLC) as the official New York City taxi zones, which is also the standard used in the record data on for-hire vehicle trips.

I automated collection of Uber data from February to September 2016. Each observation contains a surge multiplier (multiplied to fares of all trips reserved at the given timestamp and originating from the given location) and a waiting time (the estimated time of arrival for a partner driver for trips reserved at the time and location in which the query was made, rounded to the nearest positive integer).\(^5\) The average time gap between two successive observations at a given location is 11.5 minutes. Surge multipliers usually change frequently: successive observations of surge prices at the same location are rarely equal when at least one of them is larger than 1. The heat maps in Fig. 2.1 display the average surge multipliers and waiting times across all locations in NYC.

![Heat maps of average surge multiplier (left) and waiting time (right).](image)

**Figure 2.1:** Heat maps of average surge multiplier (left) and waiting time (right).

\(^5\)In all cases, UberX and UberPOOL services are subject to the same surge multiplier.
Figure 2.2: Average surge multiplier against average waiting time across all locations.

As shown in Fig. 2.1, ranks of surge multipliers and waiting times are negatively correlated: areas with highest price surges (Midtown and lower Manhattan) have the lowest waiting times. On the relation between the two metrics, studies have found empirical evidence consistent with Uber’s claim: surge multipliers are set to filter excess demand (Hall et al. 2017) and incentivize entry of drivers (Chen 2015) with the goal of ensuring short-enough waiting times for passengers. This claim does not seem to contradict the data. Across all hours, and regardless of the surge multiplier, the average waiting times for locations observed were approximately bounded below by the thresholds of 5 minutes in Manhattan and 7 minutes in other boroughs.\(^6\)

\(^6\)Locations with average waiting times longer than 8 minutes consist mainly of islands and parks.
The vacant-cab dynamic search model in chapter 1 predicts that surge pricing redistributes existing drivers across neighboring locations. All else equal, an increase in the expected payoffs at neighboring locations causes vacant drivers’ policy to put less weight on searching in the current location. Chen, Mislove, and Wilson (2015) have collected trajectories of vacant available Uber vehicles and found that, when a location has a price surge and its neighboring areas do not, these vehicles gravitate toward the surging location.

Reduced-form analysis on the collected data set on Uber waiting time and surge pricing is consistent with Chen, Mislove, and Wilson (2015). I regressed waiting times in minutes at each location in Manhattan on the average surge multiplier of its neighboring zones when the area does not have a surge (surge multiplier = 1). I constrained this regression analysis on taxi zones in Manhattan because taxi zones in other boroughs are on average larger, so the geographical density of observations is likely too low to capture variation of price surges and waiting time due to redistribution of current drivers. In the regression table below, I report results on midtown Manhattan under a 3-cluster partition (see Fig. 2.7): Midtown Center, Midtown East, and Midtown West, the three locations considered in the structural model. Neighboring zones for a location are defined as other zones in the same cluster.\footnote{The zones in the Midtown Center cluster are Flatiron, Garment District, Midtown Center, Midtown North, Midtown South, Penn Station, Times Sq, Union Square; in the Midtown East cluster, Gramercy, Kips Bay, Midtown East, Murray Hill, Stuy Town, Sutton Place, UN; in the Midtown West cluster, Clinton East, Clinton West, East Chelsea, West Chelsea}
In these regression, I regressed waiting times of a nonsurging location on the average surge multipliers for its neighbors, and I include fixed effects (location, hour, and weather) in each of the samples. Regression estimates for all regression models indicate that a location’s neighbor surge prices have a significant positive effect on its waiting time. Fig. 2.4 includes the histogram for the coefficient of average neighbor price surge for all locations in Manhattan. This is consistent with the trajectory findings of Chen, Mislove, and Wilson (2015), who indicate that drivers gravitate toward high-surging areas, decreasing the number of available vehicles and therefore increasing the waiting times for the other locations.

<table>
<thead>
<tr>
<th>Neighbor surge</th>
<th>waiting time</th>
<th>waiting time</th>
<th>waiting time</th>
<th>waiting time</th>
<th>waiting time</th>
<th>waiting time</th>
<th>waiting time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.4677**</td>
<td>1.6941**</td>
<td>1.7133**</td>
<td>1.2768**</td>
<td>2.4454**</td>
<td>0.8386**</td>
<td>1.1019**</td>
</tr>
<tr>
<td></td>
<td>(0.1292)</td>
<td>(0.0761)</td>
<td>(0.1020)</td>
<td>(0.0970)</td>
<td>(0.2176)</td>
<td>(0.1407)</td>
<td>(0.1617)</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Sample</td>
<td>Manhattan</td>
<td>Manhattan</td>
<td>Midtown</td>
<td>Midtown</td>
<td>center</td>
<td>east</td>
<td>west</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0003</td>
<td>0.5083</td>
<td>0.0152</td>
<td>0.1902</td>
<td>0.1488</td>
<td>0.2565</td>
<td>0.1546</td>
</tr>
</tbody>
</table>
Figure 2.4: Coefficients of average neighbor surge for all locations in Manhattan.

These empirical findings in Uber data suggest that flexible fare schedules can improve e-hail taxi markets. Even though taxi drivers have more rigid shift arrangements so entry effect by Uber drivers who operated under “gig-economy” shift schedules presented in Chen (2015) and Hall et al. (2015) is not plausible, the evidence presented above suggests that dynamic pricing can potentially induce short-run spatial redistribution of existing drivers to better match varying demand across location. This potential provides a basis for e-hail platform operators to employ similar policies in order to adapt to demand surges and ultimately improve passengers’ and drivers’ welfare.
2.3 The Model

2.3.1 Model Environment

There are $J$ competing e-hail dispatcher platforms and $C^j$ partner vehicles under each platform $j \in [1, J]$. Across locations and time periods $(z, t)$, the number of partner vehicles under platform $j$ and the number of riders choosing $j$ are $\{c^j_{z,t}\}$ and $\{r^j_{z,t}\}$. The total arrival of commuters at each location $z$ and time $t$ follow an exogenous process. Dispatchers act as planners to set a pricing rule $\nu^j(z, t)$ across all locations for $t \in [1, T]$. As will be discussed in the next subsection, let $Z$ be the total number of locations, a pricing rule $\nu$ is a function the assigns the fare for each of the $Z \times (Z + 1)$ trip types.\(^8\) In the cases considered in this paper, $\nu$ may be a deterministic fare multiplier vector that is determined by the pickup location, or a function that responds to instantaneous changes of supply and demand. Drivers under dispatcher $j$ observe the pricing rule $\nu^j$ and make spatial search decisions to optimize their streams of payoffs within the time frame.

At the beginning of each period, an exogenous number of commuters arrive at each location. Drivers who are vacant on each platform make search decisions and move to locations according to their policy functions. Trip fares implied by $\nu^j$ and expected waiting times are observed by commuters, who choose which dispatcher service to use, or else an outside option, to travel to their destination locations $z'$. The number of matches produced on platform $j$ is determined by the matching function $m(c, r)$. Passengers who do not get matched will take the outside option. Logistics of the model for the case $J = 2$ is illustrated in Fig. 2.5.

Passengers’ expected waiting time is determined by an exogenous function $G(c, r)$ where $c$ and $r$ correspond to the cab supply and ride demand in a given location. $G$\(^8\)With $Z$ pickup locations there are $Z + 1$ destinations including the outside region.
Figure 2.5: Logistics of the model.
is decreasing in $c$ and increasing in $r$: more available cabs decrease waiting time, and a large number of passengers choose the same platform causes the waiting time for that platform to increase.\footnote{I used a simulation model to estimate $G$ and I found that $G$ can be represented as $w = G(\frac{c}{r})$, so information on market thickness is sufficient to determine waiting times.} Note that waiting times are not observed by passengers and are revealed after the collective decision by all riders have been made.

### 2.3.2 Demand Model

The destinations $z'$ for commuters at location $z$ and period $t$ are distributed according to an exogenous probability mass function $F(z'|z,t)$. Observing the prices $p^j(z',z,t)$, expected waiting time $w^j(z,t)$ and travel time $\tau^j(z',z,t)$ for each alternative, commuters choose to take one of the e-hail dispatchers, or the outside option (metro subway) to travel to $z'$.

Passengers’ decision is modeled as a static discrete choice problem. I assume that passengers commit to their choice and do not switch alternatives within a period. If they did not get matched with their chosen ride platform $j$, they commute by metro subway. The utility for a commuter arriving at location $z$ with destination $z'$ riding metro subway is:

$$U^0(z',z,t) = \alpha_0 - \gamma w^0(z,t) - \delta \tau^0(z',z,t) + \varepsilon_0$$

The utility for riding metro subway does not have a price component because the metro subway in New York City charges a flat fare for all trips at $2.75 and I included only metro riders paying full fare, excluding those who are on passes in the estimation.\footnote{I assume that passengers on weekly or monthly passes make \textit{ex ante} decisions and do not apply to the transportation choice problem.}
The utility associated for riding with for-hire vehicle dispatcher \( j \) depends on the observed fare \( p^j(z', z, t) \), expected waiting time \( w^j(z, t) \) and travel times \( \tau^j(z', z, t) \), as well as the perceived probability of getting matched with a driver (fulfillment rate) under dispatcher \( j \), \( \psi^j(z, t) \). Therefore, the utility for using dispatcher \( j \) is given by:

\[
U^j(z', z, t) = \psi^j(z, t) \left( \alpha_1 - \beta p^j(z', z, t; \nu^j) - \gamma w^j(z, t) - \delta \tau^j(z', z, t) \right) + (1 - \psi^j(z, t)) U^0(z', z, t) + \varepsilon^j
\] (2.2)

The unobserved components \( \varepsilon^j \) are assumed to be extreme-value type 1 distributed. Waiting times \( w^j \) and matching probabilities \( \psi^j \) are endogenous and depend on the choices made by passengers in the same location. Specifically, expected waiting times and matching probabilities are both functions of the number of passengers choosing the same platform in the same location. As decisions made by other commuters at the same time and location are unobserved by each commuter, she develops beliefs on the number of passengers choosing each alternative and the implied expected waiting times and matching probabilities. As will be discussed in next subsections, these beliefs are consistent in equilibrium.

2.3.3 Decentralized Driver’s Problem

Drivers do not multi-home. Each driver belongs to exactly one dispatcher \( j \), and he does not switch.\(^{11}\) He is aware of the pricing time frame \( T \) which he optimizes his dynamic stream of payoffs observing the pricing rule \( \nu^j \). I assume that a driver commits to pick up the passenger and does not cancel once he has accepted the request.

\(^{11}\)In practice, managers of taxi fleets are required to pay a fee to the platform, making multi-homing expensive for drivers.
A driver vacant at location $z$ at period $t$ decides in which adjacent location in the neighborhood he will search for rides in that period.\textsuperscript{12} With notation on $j$ suppressed, let $V$ denote a cab driver’s value function, and $\{\lambda(z)\}_z$ be his system of belief of pickup probability, and $p, Q$ be the fare he receives (implied by pricing regime $\nu$); in addition, state transition across destination zone $z^d$ and time of dropoff $t^d$. I consider the most general form of the dynamic-search problem for the driver, in which the pricing policy $\nu$ is not revealed to the driver until collective supply and demand decisions are made. The driver develops a system belief $\tilde{\nu}$ on the pricing rule, and the Bellman equation for the dynamic programming problem is:

$$V(z, t; \lambda, \tilde{\nu}) = \max_{z'} \left\{ \lambda(z') \left( \sum_{z^d, t^d} Q(z', t, z^d, t^d) (p(z', t, z^d, t^d; \tilde{\nu}) + V(z^d, t^d; \lambda, \tilde{\nu})) \right) + (1 - \lambda(z')) V(z', t + 1; \lambda, \tilde{\nu}) + \varepsilon(z, t, z') \right\}$$  \hspace{1cm} (2.3)

$\varepsilon(z, t, z')$ is the extreme-value distributed unobserved component of the action-specific value. Drivers belonging to the same network $j$ have the same system of belief, $\lambda^j(z, t)$. The value of continuation at $t = T + 1$ is normalized to zero.

\textbf{2.3.4 Competing Dispatchers’ (Planners’) Problem}

Dispatchers act as planners for the network of partner vehicles and they set pricing rules $\nu^j$ to maximize their objective functions, denoted by $\Pi^j$. The set of domains for pricing rules is discussed in the following subsection. Operators of dispatchers set prices for each type of trips to achieve desired spatial distribution of drivers, considering responses of passengers and movement of drivers induced by their pricing policy $\nu^j$ and its competitors pricing policies $\nu^{-j}$. Here I define an equilibrium mapping $\Omega$ that maps the current number of vehicles under each platform and the current

\textsuperscript{12}To simplify the computation, I drop the assumption that a cab stays in the originating zone for $x$ minutes.
total number of commuters to equilibrium distributions of supply and demand given
pricing rules set by all dispatchers.\textsuperscript{13} Mathematically, dispatcher’s problem can be
represented by:

$$
\max_{\nu_j} \sum_{z,t} \Pi_{z,t}^j(c_{z,t}^j, r_{z,t}^j; \nu_j, \nu^{-j}) \text{ subject to } (c_{z,t}^j, r_{z,t}^j) \in \Omega(\{c_t^j\}_{j,t}, r_{z,t}; \nu_j, \nu^{-j})
$$

\textbf{2.4}

I considered the class of objective functions \(\Pi_{z,t}^j\) such that they are a weighted sum
of drivers’ and passengers’ surpluses at which \((z,t)\) pairs with weights \(\omega^j, 1 - \omega^j\).\textsuperscript{14}
The average consumer surplus at \((z,t)\) is defined by the expected utility of riding \(j\)
with respect to the outside option, multiplied by \(\mu^j\): the probability of choosing \(j\) and
integrating out the distribution of destination locations \(z'\):

\[\Pi_{z,t}^j(\nu_j, \nu^{-j}) = \omega^j m^j(c_{z,t}^j, r_{z,t}^j) \cdot p(z, t; \nu_j) + (1 - \omega^j) \Xi_{z,t}^j(\nu_j, \nu^{-j})\]

\[\Xi_{z,t}^j(\nu_j, \nu^{-j}) = \int_{z'} \mu^j(z', z, t; \nu_j, \nu^{-j}) ((U^j - U^0)(z', z, t; \nu_j, \nu^{-j})) F(dz'|z, t)\]

The platform’s problem takes equilibrium response from partner drivers and pas-
ssengers as given. When \(J > 1\), the platform also considers the fare charged by com-
peting platforms, \(\nu^{-j}\) and the best response function for dispatcher \(j\) is: \(\nu^j*(\nu^{-j}) = \max_{\nu_j} \sum_{z,t} \Pi_{z,t}^j(\nu_j, \nu^{-j})\).

\textbf{2.3.5 Dispatchers’ Pricing Problem}

In a setup with \(Z\) locations, the complete pricing regime space (action space for dis-
patchers) is \(\mathbb{R}^{Z \times (Z + 1)}\), in which the dispatcher sets the fare rate for each possible
trip transition. The state space for platforms include current spatial distribution of
cabs and riders, and the fares charged by competing platforms. As the number of

\textsuperscript{13}When there are multiple equilibria, \(\Omega\) implicitly applies the equilibrium selection rule
introduced in the following section.

\textsuperscript{14}Other candidates include minimum deviation from a maximum waiting time threshold,
maximum demand fulfillment rate and number of matches produced.
dispatchers increases, the dimension of $\nu^{-j}$, the strategies of competing platforms increases exponentially. I constrained the pricing analysis of this paper into 2 pricing schemes that can be described by low dimensional $\nu$s: pre-determined regional multiplier and real time linear surge multiplier.

Under both pricing schedules, fare of a trip is calculated as the product of base fare and a multiplier. The base fare of a trip is calculated as the meter fare, which I approximated as the sum of the minimum payment, linear distance and time fares. Under the spatial pricing scheme, the dispatcher varies prices by choosing a multiplier $\nu_z$ for all trips originating from location $z$, so that passengers in $z$ will pay $\nu_z^j \geq 1$ times of the meter fare. I assumed that the regional multipliers are known by drivers and are time-invariant within the time frame of the model. In a model with $Z$ locations, the pricing policy domain for a dispatcher is $\nu \in \mathbb{R}^Z$ and its state space is $\nu^{-j} \in \mathbb{R}^{Z \times (J-1)}$.

The second dynamic pricing regime is a linear real-time surge multiplier schedule. Under this mode, the surge multiplier responds to real-time market tightness. This is the common pricing schedule adopted by peer-to-peer ride-sharing platforms (surge pricing for Uber and prime time for Lyft); fares are updated in real time. I use a linear function to approximate a real-time algorithm. Formally, when there are $c_{z,t}^j$ vehicles and $r_{z,t}^j$ riders that request rides on platform $j$ at $(z, t)$, the surge multiplier is:

$$
\nu_z^j(c_{z,t}^j, r_{z,t}^j) = 1 + \theta^j \left[ \frac{r_{z,t}^j}{c_{z,t}^j} - 1 \right]^+, \quad \theta^j > 0
$$

When there is excess demand, the rider-to-vehicle ratio is larger than 1 and this causes the surge multiplier $\nu_z$ to exceed 1 and triggers surge pricing. This is consistent with Uber’s claim that surge pricing is to clear the market when there is excess demand. I constrained that the sensitivity parameter to be the same across all locations, and prices do not vary within the 10-minute period. The parameter $\theta^j$ can be thought of as the price sensitivity to excess demand. A zero value of $\theta$ corresponds to
the case of no surge pricing, and when there are twice as many riders as drivers, the surge multiplier would be $1 + \theta$.

Platform compete in a Bertrand price competition, as dispatchers set prices and the quantities are decided subsequently by passengers. The Bertrand-Nash equilibrium under this setup is formally defined in the subsection below.

2.4 Competitive Equilibrium

In chapter 1, I presented a competitive equilibrium in which drivers optimize their decisions conditional on rational beliefs about pickup probabilities. Drivers’ search decisions interact among themselves under the same matching platform: given net demand for rides under the same dispatcher, more cabs in the same location leads to more competition and therefore a lower pickup probability. With the independent-draw e-hail matching function and street hail Cobb-Douglas function, the first chapter defined a competitive equilibrium and developed a belief iterative method to solve for them.

In this chapter, I accounted for response from demand side and subsequent equilibrium “coordination” within and across both sides of the market to define a competitive equilibrium in a broader sense. Specifically, the equilibrium covers the following strategic considerations for drivers and riders. When a driver makes search decisions he acknowledges the facts that 1. decisions made by other drivers determine pickup probabilities (peer competition) and 2. passengers’ waiting times change in respond to the resulting spatial cab distributions from drivers’ decisions, which ultimately determines the demand for cabs (under the driver’s dispatcher). Likewise, a passenger makes her transportation choice considering that 1. excess demand for cabs under a dispatcher increases waiting times, triggers surge pricing (when appli-
cable) and decreases request fulfillment probabilities (peer competition), but 2. excess demand attracts drivers. Taking these strategic interactions into account, agents are optimizing subject to their rational expectations of beliefs in equilibrium. The formal definition is presented below.

**Definition:** A competitive equilibrium in this model is \((S_0, \{\tilde{r}^j\}, \{\lambda^j\}, \{\tilde{\nu}^j\}, \{V^j\}, \{\phi^j\}, \{G\}, \{m^j\})\): initial states \(S_0\), a system of beliefs on ride demand, a system of beliefs on pickup probabilities \(\lambda^j\) and pricing policies set by dispatcher, \(\tilde{\nu}^j\), value and policy functions \((V^j, \phi^j)\), and systems of waiting time and matching functions \(G^j, m^j\) such that:

- The value and policy functions \((V^j, \phi^j)\) solve drivers’ optimization problem given their beliefs on pickup probabilities \(\lambda^j\) and dispatcher-set pricing policies \(\tilde{\nu}^j\).

- The system of belief on ride demand is consistent with the demand given spatial cab distribution \(c^j(\phi; S_0)\) and beliefs on waiting times \(w^j(\tilde{r}^j)\), matching probability \(\frac{m^j(c^j, \tilde{r}^j)}{\tilde{r}^j}\).

- The system of belief on pickup probabilities is rational and consistent with the matching function and equilibrium demand: \(\lambda = \frac{m(c(\phi; S_0), r(\tilde{\nu}, w))}{c(\phi; S_0)}\).

- The system of belief on pricing policies is rational and consistent with equilibrium cab supply and ride demand under pricing rules set by dispatchers: \(\tilde{\nu} = \nu(c(\phi; S_0), r(\tilde{\nu}, w))\).

Taking demand as given, the rationale of supply side equilibrium is equivalent to the competitive equilibrium definition in chapter 1. At equilibrium, drivers are optimizing while having rational expectation. Likewise, this equilibrium definition is an
“oblivious equilibrium”. Instead of conditioning on other agents’ strategies, the state variables on which agents make decisions are belief vectors that internalize competition from peers and supply/demand conditions. In particular, the belief of pickup probabilities on platform \( j \) given by \( \lambda_j \) is sufficient for the information on competition from drivers across and within platforms. On the other hand, at equilibrium, the belief of expected waiting time \( \tilde{w}^j \) reflects the actualized waiting time caused by drivers’ search decisions and choice of other passengers in the same location. Given rational expectation in passengers’ beliefs, and the flexibility for drivers to take into account passengers’ responses to spatial distributions of cabs induced by their decisions. Notice that the set of equilibria presented in chapter 1 is a subset of the competitive equilibria defined above, with the restriction that the passengers’ responses to waiting times is inelastic. Taking the demand utility, waiting time and matching functions for street hail taxis estimated in chapter 1, the competitive equilibrium solution under this extended definition predicts similar levels of pickups in midtown Manhattan (at 5 p.m.) in the for-hire vehicle market, pickup probabilities (72%) with a average lower waiting time,\(^{15}\) suggesting that passengers under the extended model are behaving more optimally and are able to “coordinate” among themselves to utilize existing resource of taxicabs, so unmatched demand is reduced. Equilibrium criteria in chapter 1 are sufficient to model drivers’ search behavior, but the flexibility on passengers’ choices under this definition allow room to model optimal behavior from the demand side too. As the for-hire vehicle industry is now operated by dispatcher platforms, passengers are more well-informed about waiting times and vehicle avail-

\(^{15}\)This comparison is only an approximation to illustrate the relationship between the sets of equilibria under two frameworks. There are settings in the model presented in chapter 1 that this model does not account for (e.g. in chapter 1, drivers stay in an area for \( x \leq 10 \) minutes for migrating to another zone, and the model in this chapter treats midtown Manhattan as a closed system.
ability. Therefore, extending the demand side in the equilibrium concept allows us to make more robust predictions under various regimes.

The restriction of rational beliefs about pricing policies \( \nu \) is always fulfilled if \( \nu \) is determined \textit{ex ante} to drivers’ and passengers’ decisions. This is the case in spatial pricing in which the regional multiplier is known to drivers and passengers across all regions. Suppressing real time variations in \( \nu \), The figure below illustrates the structure of an equilibrium. Drivers develop a system of belief about pickup probabilities, \( \lambda \), resulting in the policy function \( \phi \) which is a solution for their dynamic programming problem. \( \phi \) induces a spatial distribution of cabs which passengers take as exogenous. Passengers coordinate and develop rational expectations on waiting times, which aggregated to actualized demand \( r \) and pickup probability \( \hat{\lambda} \). At equilibrium, \( \lambda \) converges to \( \hat{\lambda} \). The next section covers an algorithm solving for the inner loop on the demand side.

![Figure 2.6: Structure of competitive equilibrium.](image)
2.4.1 Existence and Uniqueness

The existence of equilibria follows from the proof presented in chapter 1, but now drivers need to consider feedback of their actions from the demand side. I considered the defined competitive equilibrium as a nested problem and I presented an informal proof for its existence in two steps. I first show that equilibrium on the demand side exists for any spatial distribution of drivers with positive number of vehicles in all locations. Then, given this continuous mapping from spatial distribution of drivers to demand response, I showed that equilibrium exists on the supply side.

Given drivers’ spatial distribution, the demand equilibrium exists if and only if passengers’ beliefs on demand for each platform $j$, $\tilde{r}_j$ and the resulting price given $\nu^j$ are rational. When $\nu^j$ is determined by the pickup location (spatial pricing), the second condition will be fulfilled. Here I considered the case of surge pricing, so $\nu^j$ is endogenous. Mathematically, these two conditions are expressed as:

$$\tilde{r}_j = \frac{\int_{z'} \mu^j(z', z, t; \tilde{\nu}_j, \tilde{\nu}^{-j}) F(dz'|z, t)}{1 + \theta \left( \frac{\tilde{r}_j}{c^j} - 1 \right)^+}$$

$$\nu^j = G(c^j, \tilde{r}_j)$$

In other words, a demand side equilibrium is a fixed point of the mapping that maps the beliefs of platform demand to expected number of passengers choosing $j$. Note that this mapping is continuous. Because utility shocks for passengers are extreme-value distributed, choice probability $\mu^j$ is a logit function of the utilities which is continuous in the utility term $U$. If $G$ is continuous and $c^j > 0$, waiting time $w^j$ and pricing policy $\tilde{\nu}^j$ are continuous in $\tilde{r}_j$. As the domain and support for the mapping is bounded in $[0, r_{z,t}]^J$, the mapping from platform demand belief to expected number
of passengers is a continuous function that maps from $[0, r_z]$ onto itself. Therefore, Brouwer fixed point theorem guarantees the existence of a fixed point.

Denoting the equilibrium demand given a distribution of drivers $c$ by $\{r^j(c)\}_j$, equilibrium on the supply side requires that drivers’ beliefs on pickup probabilities $\lambda$ and pricing policy $\tilde{\nu}$ be consistent with the matching function and the equilibrium demand induced by drivers’ dynamic search policy solution for the dynamic programming problem for drivers given $(\lambda, \tilde{\nu})$. That is,

$$
\lambda^j = \frac{m\left(c^j(\lambda^j, \tilde{\nu}^j), r^j(c(\lambda, \tilde{\nu}^j))\right)}{c^j(\lambda^j, \tilde{\nu}^j)} \tag{2.10}
$$

$$
\tilde{\nu}^j = 1 + \theta \left[\frac{r^j(c(\lambda, \tilde{\nu}^j))}{c^j(\lambda^j, \tilde{\nu}^j)} - 1\right]^+ \tag{2.11}
$$

It follows that a supply-side equilibrium is a fixed point for the mapping of beliefs of pickup probabilities and pricing policies to their actualized counterparts. When we invoke the Brouwer fixed-point theorem and continuity of matching function and pricing functions, it suffices to show that the spatial distribution of drivers is continuous in $\lambda^j, \tilde{\nu}^j$ and that $\tilde{\nu}^j$ is bounded. Continuity of $c$ in $\lambda$ is presented in chapter 1, and continuity of $c$ in $\tilde{\nu}$ is derived from the fact that the value function is linear in $\nu$ and the policy function takes the logit form. Next, note that, because of the extreme-value shock in the alternative specific-value function, the driver distribution is positive across all locations.\textsuperscript{16}

A large number of simulations supports the conjecture that the demand-side equilibrium is unique up to the initial spatial distribution of drivers. Simulation results are also consistent with uniqueness when there is just one dispatcher. When the market is a duopoly, however, supply-side equilibrium is not unique. It turns out that competitive-equilibrium outcomes depend on the starting beliefs of the drivers.\textsuperscript{16}

\textsuperscript{16} Small values of $c$ would make convergence in computation slow, but in practice these values are sufficiently large to guarantee convergence in a small number of iterations.
pointed out by Lee (2017), shocks that produce spatial distributions randomize equilibrium outcomes for drivers and platforms. I impose the equilibrium selection rule in which I select the initial beliefs of drivers such that platforms with the same numbers of vehicles and the same pricing rules end up getting the same expected demand for rides across all locations.\(^\text{17}\)

2.4.2 Computation Algorithm

I introduced an iterative algorithm to solve for a competitive equilibrium defined above. The spirit of the algorithm is similar to the simulated fixed point iteration on beliefs, with the main difference that waiting times, demand and supply are evaluated according to the analytic forms of their expected values, rather than approximated by simulated averages. In other words, a “mean field” / “continuum agent” approximation is imposed to smooth out the beliefs about waiting times and pickup probabilities that would otherwise be discontinuous due to the discreteness of (finite) supply and demand. In chapter 1, I overcame this challenge by approximating expected beliefs with multiple simulations. Because this model is smaller and simplifying assumptions were made,\(^\text{18}\) analytic forms of beliefs which are continuous and converges rapidly can be computed. Second, a nested “polyalgorithm” is required for solve general equilibria in this model because it requires simultaneous convergence of drivers’ and passengers’ beliefs. I decomposed the equilibrium problem into a nested problem as follows: given any spatial distribution of drivers, passengers develop rational expectations of waiting times and fare rates to make transportation choices. Once passengers learn to coordinate and optimize under systems of rational beliefs, the resulting equilibrium demand is revealed to drivers which they consider a function of their decisions.

\(^{17}\)There are multiple equilibria in some cases, and I selected by setting initial beliefs with values 1 at every state.

\(^{18}\)e.g. I ignored the time spend on road in the originating location.
Therefore it suffices to develop an algorithm for drivers to learn beliefs on pickup probabilities, taking into account equilibrium response in the inner loop as a function of supply side decisions. The nested algorithm puts drivers’ iteration in the outer loop because drivers’ problem is dynamic and passengers’ problem is static, drivers’ beliefs are updated every episode whereas passengers’ beliefs are updated within a period. To sum up, the equilibrium algorithm is a nested belief iteration, in which passengers’ beliefs (inner loop) are iterated until convergence for every time period in every iteration of drivers’ belief (outer loop). Equilibrium is achieved when drivers’ beliefs converged to the actualized pickup probabilities.

Updates of beliefs are implemented through relaxation methods: updated beliefs are weighted sums of last iteration’s belief and the actualized variable in the current iteration. In chapter 1, the relaxation constant is taken as 1; in this model, a biconvergence requirement and endogenous response from passengers render the realized equilibrium’s supply and demand more volatile, and the relaxation constant $\eta$ needs to be smaller than 1. As will be illustrated in Appendix D, convergence of the algorithm builds on the contraction-mapping property of the logit function.
Algorithm 4: Competitive Equilibrium Algorithm (given $\nu^j$).

1: Initialize system of beliefs $\lambda^j$, $\tilde{\nu}^j$ for drivers on platform $j \in [1, J]$.  
2: while non-convergence of drivers’ beliefs do  
3:   $\forall j \in [1, J]$, solve for dynamic programming optimal policies $\phi^j(z, t)$.  
4:   for $t = 1, 2, \ldots, T$ do  
5:      Initialize passengers’ beliefs on demand $\tilde{r}^j$ for all $j$.  
6:      while non-convergence of passengers’ beliefs do  
7:         Given cab distribution and belief on demand, compute expected waiting time and fare rates.  
8:         Evaluate passengers’ choice probabilities to compute implied demand.  
9:         With implied demand $\tilde{r}^j$, update belief $\tilde{r}^j \leftarrow \eta \tilde{r}^j + (1 - \eta)\hat{r}^j$. Evaluate convergence criterion $\|\tilde{r} - \hat{r}\|$.  
10:        Go back to step 7 and continue to convergence.  
11:      end while  
12:      Using equilibrium demand and policies of drivers $\phi^j$, compute implied matches, pickup probabilities and prices $\hat{\lambda}^j, \hat{\nu}^j$.  
13:   end for  
14:   Update drivers’ beliefs $\hat{\lambda}^j(z, t) \leftarrow \eta \hat{\lambda}^j(z, t) + (1 - \eta)\tilde{\lambda}^j(z, t)$ and $\hat{\nu}^j(z, t) \leftarrow \eta \hat{\nu}^j(z, t) + (1 - \eta)\tilde{\nu}^j(z, t)$ for all $(z, t)$.  
15:   Evaluate convergence criterion for drivers’ beliefs.  
16:   Go back to step 3 and continue to convergence.  
17: end while

2.5 Computation and Calibration

To apply this model of the for-hire vehicle industry in an empirical setup, I consider a 3-location model based on Manhattan Midtown (north of 14th St. and south of 59th St.). As shown in Fig. 2.7, I partition Manhattan Midtown into 3 zones: Midtown Center, Midtown East, and Midtown West. As for the model’s time frame, I focus on 5 p.m. to 5:50 p.m. For this model, I have chosen this setup of place and time for two reasons. First, Midtown can be approximated as a closed system. It is a large area in which trips have high probabilities of dropoffs ending up in one of the locations. From 5 p.m. to 5:50 p.m., about 70% of all trips picked up in Midtown result in dropoffs in one of the Midtown locations. In this model, I treat Midtown as a closed system and
consider entry of vehicles from outside the system only through an exogenous process
of dropoffs originating from the outer zone. Furthermore, Midtown is the area with
most price surges in the Uber pricing dataset, reflecting a tight rider-cab ratio. This
excess demand is magnified from 5 p.m. to 5:50 p.m. As most taxi cab drivers have
shift-changing times around 6 p.m., day shift drivers have to return to the garage to
hand vehicles over to night shift drivers, creating a shortage of cab supply. Therefore,
Midtown at 5 p.m. is ideal, since the goal of the exercise is to study the welfare impact
of dynamic pricing schedules under excess demand.

![Figure 2.7: Three-location setup in the model.](image)

2.5.1 Model Environment Variables

In counterfactual experiments, I assume that e-hail taxi market is a monopoly or
duopoly, and included Uber as an option available to passengers. The outside option
is metro subway. The initial number of taxicabs at the beginning of the model is
calibrated with the prediction of cab numbers in chapter 1. At each period, there
is an exogenous number of drop offs from outside midtown, inferred directly from
the TLC trip record data set. The number of passengers at each location is taken as the sum of taxi and Uber passengers. Fig. 2.8 shows the arrival of passengers and number of available cabs (without considering matching) across three locations. Midtown Center and East are busy districts and they have similar levels of passenger arrivals, while midtown West has significantly lower passenger arrival rates. It can be seen that even without considering matched cabs there is excess demand in Midtown Center and Midtown East.

![Passenger and cab arrivals at each location.](image)

**Figure 2.8:** Passenger and cab arrivals at each location.
The action space for drivers consists of staying in the current zone and searching in the adjacent zone(s). Therefore, feasible searching zones for Midtown Center include all three locations, whereas vacant drivers in Midtown East and Midtown West have only two choices: to stay or to search in Midtown Center. For each of the locations, trip pickups can result in a dropoff at either one of the three locations in the model or out of midtown, in which case the value of continuation is normalized to zero. Travel times for each type of trip are inferred from taxicab trip record data. State transition function for cab drivers’ dynamic programming problem is inferred from data, where I partition destinations into the three zones in the model plus a collective destination including dropoff locations outside the model. As in the case presented in chapter 1, trip duration for a given origin-destination pair is stochastic, but most trips end after one 10-minute period. I compute the net-of-multiplier meter fare for taxicab services by means of the TLC meter rule, and I apply a 15% tip percentage.
I treated Uber as an exogenous platform, and I took the surge multipliers and waiting times at each location and time as the observed average in the Uber API data set. The waiting and travel times for metro (outside option) are estimated by Google Maps. Waiting time for metro trips include average time spent walking to the metro stations.

2.5.2 Demand Parameters

I take the demand estimates from chapter 1 to account for passengers’ responses to prices and waiting times offered by each mode of transportation. The demand model presented in chapter 1 is a nested logit model, but evaluating nested choices is computationally expensive. Therefore, I make the simplification assumption of using a multinomial logit model, assuming common coefficients on prices and waiting times. As documented in chapter 1, the demand model is estimated by means of a method of simulated moments to match the market shares of taxicabs, Uber, and metro across locations and time periods.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Point Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>Base utility for outside option</td>
<td>$-22.25$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>Base utility for ride</td>
<td>$10.75$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Coefficient on prices</td>
<td>$-12.00$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Coefficient on waiting time</td>
<td>$-7.783$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Coefficient on travel time</td>
<td>$-2.873$</td>
</tr>
</tbody>
</table>

2.5.3 Waiting Time and Matching Functions

I estimated the function for waiting time by simulating both searching e-hail drivers and waiting passengers on an approximation of midtown Manhattan on a rectangular
lattice, using the regular grid structure of block structures in Manhattan (Frechette, Lizzeri, and Salz 2017); details of simulation procedures are included in Appendix F. I simulated the arrival process of passengers in the points on the grid. Within a location, heterogeneity in passenger arrival rates is approximated by the relative density of pickups observed in taxicab trip record data. Drivers move around the location according to a “heat map rule” and gravitate toward areas that are dense in demand. Passengers leave if their waiting times exceed 10 minutes, and drivers leave the area with an exogenous probability calibrated by the estimates of the drivers’ value function in chapter 1.

Unlike street-hail matching, a match does not require physical contact of passengers and drivers. Instead, when a request is made, the dispatcher platform searches for the nearest driver in terms of block distance. A request’s waiting time is the time elapsed until the platform finds a match plus the time required for the nearest driver to travel the block distance from the passenger.\(^{19}\) Fig. 2.10 shows the simulated waiting time for various numbers of cabs in Midtown East against the rider-driver ratio \(r_c\). According to simulation results, the average waiting time is a function of \(r_c\), the rider-driver ratio.

Once rider-driver ratio exceeds 1, waiting time increases steeply, and linearly at a lower rate when the ratio is larger than 1.5. Therefore, I use a sum of linear and logistic functions to fit the waiting-time function and the functional form; the non-linear least-squares parameter estimates for the locations are:

\(^{19}\)As pointed out by Feng, Kong, and Wang (2017), waiting time also depends on relative directions of traffic (clockwise or anticlockwise) at the positions of drivers and passengers. There is, too, a matching time on the platform’s side. I account for these factors by adding a 2-minute compensation term.
Figure 2.10: Simulated and fitted waiting times.
\[ w = \rho_0 + \rho_1 \frac{r}{c} + \rho_2 \frac{\exp \left( \rho_3 \left( \frac{r}{c} - \rho_4 \right) \right)}{1 + \exp \left( \rho_3 \left( \frac{r}{c} - \rho_4 \right) \right)} \] 

(2.12)

<table>
<thead>
<tr>
<th>Table 2.3: Estimates of waiting time function.</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>(\rho_0)</td>
</tr>
<tr>
<td>(\rho_1)</td>
</tr>
<tr>
<td>(\rho_2)</td>
</tr>
<tr>
<td>(\rho_3)</td>
</tr>
<tr>
<td>(\rho_4)</td>
</tr>
</tbody>
</table>

The e-hail matching function introduced in chapter 1 embodies friction due to small numbers of e-hail drivers on a platform. The e-hail matching function presented in chapter 1 is: \( m = \min\{c, r\}(1 - \xi^c) \) for some constant \( \xi < 1 \). Using the e-hail rule set by TLC that e-hail matches that take place in Manhattan midtown and downtown must be within 0.5 miles, I used the waiting time simulation to calibrate the value of \( \xi \). According to this function, a large number of cabs increases the expected matches produced, and also the probability that the frictionless number of matches were produced. Therefore, in this model, welfare changes arise from variation in prices, waiting times and probability of getting matched due to allocation of drivers and the size of driver network under the platform.

2.5.4 Equilibrium Computation

Given a pricing policy vector \( (\nu_1, \ldots, \nu_J) \) I used algorithm 4 to solve for competitive equilibrium including demand side responses on transportation choices, supply side search decisions and resulting matches. The relaxation parameters for drivers’ and

\(^{20}\hat{\xi} = 0.9912\)
passengers’ beliefs are set as 0.2 and 0.005 respectively. Compared to the support of drivers’ beliefs on pickup probabilities in \([0, 1]^{Z \times T}\), support of passengers’ beliefs for equilibrium demand is \([0, r_{z,t}]^T\) where \(r_{z,t}\) is the total number of passengers in \((z, t)\) (Fig. 2.11). Small deviations of passengers’ beliefs can lead to large fluctuations in responses, as magnitude of \(r_{z,t}\) can be as large as 950. Therefore, I took a small relaxation constant for passengers’ loop and this required a large number of iterations to converge.

As discussed in the previous section, uniqueness of equilibrium under a given pricing rule \(\{\nu^j\}\) is not guaranteed when there are more than 1 dispatcher. A large number of simulations suggests that uniqueness is up to the starting beliefs on pickup probability for drivers. To standardize the analysis on pricing experiments, I focus on the class of equilibria with the condition that, if \(\nu_1 = \nu_2\) then the demand ratio of platform 1 to 2 is \(c_1\) to \(c_2\), the ratio of the number of drivers under each of them. This class of equilibria can be achieved by setting initial identical beliefs of drivers across platforms and iterate over the algorithm introduced above.\(^{21}\)

\(^{21}\)The same set of equilibria can be achieved by setting initial beliefs for platform \(j\) to be its market share.
This section discusses dynamic pricing experiments under monopolistic and duopolistic e-hail taxi markets. The benchmark results (equilibrium under meter fare/no dynamic pricing) were presented in Table 2.4.\textsuperscript{22} The benchmark results for duopoly presented are estimated from a symmetric duopoly market. In this case, each of the two platforms partners with 50% of the drivers. The monetized values of consumer surplus is calculated by the total excess utility of riding e-hail cabs with respect to the outside option for all commuters (see definition of $\Xi$ in 2.6), normalized to monetary units using the price coefficient in the utility function.

The comparison of monopolistic and duopolistic cases demonstrates network externalities. The monopolistic case exhibits higher matching efficiency: total matches is more than double the symmetric duopoly case, drivers’ pickup probability and fulfillment rates for requests are both higher and the total consumer surplus under the two platforms in duopoly is lower than the monopolistic case. This echoes the result presented in chapter 1: comparing with a fragmented e-hail market, a universal e-hail taxicab dispatcher provides more gains for consumers.

Supply and demand at the three locations for the monopolistic e-hail taxi dispatcher are illustrated in Fig 2.12. Under equilibrium, there are excess demands in Midtown Center and Midtown East, and excess supply in Midtown West. Without dynamic pricing, decentralized decisions of taxi drivers create search friction.\textsuperscript{23} At each period, I consider the following metric as the measurement of search friction on platform $j$: $\Psi_j^t = \min\{\sum_z (c_{z,t}^j - r_{z,t}^j)^+, \sum_z (c_{z,t}^j - r_{z,t}^j)^-\}$. This expression can be

\textsuperscript{22}I considered e-hail without dynamic pricing the benchmark case. The transition from street hail to e-hail is discussed in chapter 1 and the goal of this chapter is to discuss welfare implications of dynamic pricing.

\textsuperscript{23}Search friction in the taxicab market is defined by Lagos (2000) as existence of excess supply and excess demand.
interpreted as the hypothetical extra matches produced when the excess supply is
allocated at the locations for which there is excess demand under a perfect matching
function. The numbers reported in table 2.4 are the Ψ summed across all periods.
As characterized by Ψ, there is more search friction under duopoly, since the total Ψ
in duopoly case is higher than in monopoly.

Table 2.4: Equilibrium for benchmark cases.

<table>
<thead>
<tr>
<th></th>
<th>Monopoly</th>
<th>Duopoly (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand</td>
<td>6,914</td>
<td>7,048</td>
</tr>
<tr>
<td>Matches</td>
<td>6,624</td>
<td>6,316</td>
</tr>
<tr>
<td>Pickup prob</td>
<td>0.9434</td>
<td>0.8435</td>
</tr>
<tr>
<td>Fulfillment rate</td>
<td>0.9580</td>
<td>0.8963</td>
</tr>
<tr>
<td>Waiting time</td>
<td>4.8420</td>
<td>4.3770</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$41,847</td>
<td>$39,400</td>
</tr>
<tr>
<td>Platform earnings</td>
<td>$79,799</td>
<td>$75,758</td>
</tr>
<tr>
<td>Ψ</td>
<td>83.26</td>
<td>93.76</td>
</tr>
</tbody>
</table>

24It is noted that Ψ is a naïve measurement because it excludes equilibrium response when
the excess cabs are actually allocated to these areas with excess demand.
2.6.1 Dynamic Pricing Experiments in a Monopolistic E-hail Market

I considered a dynamic-pricing experiment in the e-hail taxicab market when the market is operated by a monopolistic dispatching platform. Passengers choose from three choices: e-hail taxicabs operated by the monopolist platform, Uber, and metro subway. I consider two dynamic pricing schemes: fixed regional multiplier and real-time surge multiplier. In the first case, all trip fares originating from zone $z$ are
multiplied by the zone-specific multipliers \( \nu_z \), which are fixed across the time frame of the model and are revealed to drivers \textit{ex ante} to their search decisions. In the second case, I consider a dynamic pricing scheme that responds to the real-time ratio of demand to supply. When there is excess demand, surge pricing is triggered and fares are subject to a surge multiplier calculated according to equation 2.7, where I constrain the parameter \( \theta \) to lie between 0 (no surge) and 10. Under each of the pricing schemes, I consider two sets of objective functions for the monopolist platform: profit maximization and total social welfare, with equal weights assigned to drivers and passengers.

<table>
<thead>
<tr>
<th>Table 2.5: Optimal monopolist pricing.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Profit Maximizer</td>
</tr>
<tr>
<td>Center multiplier</td>
</tr>
<tr>
<td>East multiplier</td>
</tr>
<tr>
<td>West multiplier</td>
</tr>
<tr>
<td>Surge sensitivity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2.6: Monopolist equilibrium with optimal dynamic pricing.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>spatial pricing                  surge pricing</td>
</tr>
<tr>
<td>benchmark profit welfare profit &amp; welfare</td>
</tr>
<tr>
<td>Demand</td>
</tr>
<tr>
<td>Matches</td>
</tr>
<tr>
<td>Pickup prob</td>
</tr>
<tr>
<td>Fulfillment rate</td>
</tr>
<tr>
<td>Waiting time</td>
</tr>
<tr>
<td>Consumer surplus</td>
</tr>
<tr>
<td>Platform earnings</td>
</tr>
<tr>
<td>Mean price surge</td>
</tr>
<tr>
<td>( \Psi )</td>
</tr>
</tbody>
</table>

I present the optimal monopolist pricing strategies and equilibrium outcomes under each of the cases in Table 2.5 and Table 2.6. The equilibrium supply and
demand across locations and times, under each optimal pricing scheme as well as in
the benchmark case, are shown in Fig. 2.13. As one would expect, the optimal regional
multipliers are higher in the profit-maximizing case than the welfare-maximizing case.
In all cases, the multiplier for Midtown West is the lowest relative to the other loca-
tions so as to disincentivize drivers’ entry and allocate the excess supply in Midtown
West under the benchmark case to Midtown Center and Midtown East where there is
excess demand, as shown in Fig. 2.13. With the profit-maximizing objective function,
regional fare multipliers are set to induce excess supply across all locations and time
periods. Such is the case with the lowest search friction metric $\Psi$: excess demand is
mostly non-existent and very low when it is present, which also causes the low waiting
times and high fulfillment rates seen in Fig. 2.14. Demand in Midtown Center and
Midtown West decrease in response to the fare rates, but increased entry of drivers in
Midtown East generates a higher fulfillment rate and lower waiting time that trade off
against the higher fare rate, increasing demand. Nonetheless, the increase in trip fare
outweighs these gains and there is thus a loss in consumer surplus across all locations.
It is also noted that drivers’ utilization rates are lower than the benchmark, and their
earnings distribution becomes more skewed as fewer drivers get matched for a larger
total fare revenue. A similar pattern of regional multipliers is found in the welfare-
maximizing solution, where the dispatcher reallocates excess supply in Midtown West
to Midtown Center and Midtown East, and also sets slightly lower fare multipliers to
bring about mutual gains for the platform and the passengers.

The maximizing value of real-time surge parameter $\theta$ are 10 for both profit-
maximizing and welfare-maximizing platforms. To check whether welfare optimality
was driven by the gains from the platform, I considered an objective function maxi-

\footnote{I ran the model with a smaller fleet (60% of the taxicab number) to mimic Uber, and
found that a value of $\theta = 10$ would imply values of surge multipliers in equilibrium which
best matched Uber surge multipliers observed.}
mizing consumer surplus and found that the consumer-surplus-maximizing \( \theta \) was also 10. A surge-pricing parameter of \( \theta = 10 \) implies that, if the number of riders is double the number of drivers, fares will be subjected to a \( 11 \times \) multiplier. Despite this massive effect, surge pricing is rarely triggered under equilibrium. Passengers realize that requesting more than the number of available cabs will trigger surge pricing, and so under equilibrium they “coordinate” to match the supply. The supply and demand shown in Fig. 2.13 closely match each other, consistent with Hall et al. have shown (2015): real-time surge pricing filters excess demand in Midtown Center and Midtown East that would have existed without surge pricing (see Fig. 2.12). Drivers are reallocated to Midtown East by their rational expectation of demand, which also implies the probability of getting a price surge.\(^{26}\) Compared with benchmark and regional pricing cases, surge pricing yields the highest matching probabilities for drivers and passengers, implying a higher utilization rate for resources. Compared with fixed regional pricing, a real-time surge-pricing schedule also provides more flexibility for the dispatcher to respond to irregular demand patterns within a location to allocate cabs to riders. It also generates a higher total social surplus.

On top of an allocation perspective, it is also acknowledged that surge pricing is more desirable because it does not require that the operator of the dispatcher has information on innate demand patterns across locations (Riquelme, Banerjee and Johari, 2015) as in the case of spatial pricing. Also, it is more resilient to unexpected shocks (Hall et al., 2015). Off-equilibrium allocation is out of the scope of this paper, but the data and model available can allow for predictions. A simulation of

\(^{26}\)Besides fixed market thickness and fare rate, drivers are also drawn to locations with larger networks of drivers, locations that imply a higher pickup probability (see matching function). Specifically in Fig. 2.14 there is a decrease in the fulfillment rate of Midtown West, 17:30 under all-dynamic pricing, because of the smaller number of drivers in Midtown West.
off-equilibrium supply and partial equilibrium demand consistent with the supply pattern under an unexpected shock is included in the appendix G. When there demand shocks, surge pricing is triggered to filter demand, leading to higher request fulfillment rates and shorter waiting times for matched passengers.

2.6.2 Dynamic Pricing Experiments under Duopoly

Under a duopoly e-hail taxi market, I assumed that dispatchers compete in a Bertrand price competition: each dispatcher responds only to prices set by its opponents.\[27\] Formally, Bertrand equilibrium in the for-hire vehicle market is defined as:

**Definition:** A pair of strategy \((\nu^1, \nu^2)\) is a Nash Equilibrium if \(\forall i \in \{1, 2\}, \nu^i\) is the best response for \(i\) given \(\nu^j, j \neq i\):

\[
\nu^i \in \arg\max_{\nu^i} \left\{ \sum_{z,t} \Pi^j_{z,t}(\nu^i, \nu^j); . \right\}
\tag{2.13}
\]

In this section, I considered competition between platforms under spatial pricing and real time surge pricing. In each case, I computed the equilibrium across various market shares.

**Surge Pricing**

When the firms compete in the surge pricing space \(\theta\), it is found that the best responses to any value of \(\theta^{-i}\) is either 0 or 10.\[28\] It follows that under real time surge pricing schedules, duopoly competition decomposes to a normal form game with pure strategy space \(\theta^i \in \{0, 10\}\). Denote \(ns\) and \(s\) for “no surge” and “surge”, Table 2.7 presents

\[27\] Because waiting times and request fulfillment rates are determined under competitive equilibrium in e-hail markets given the fare rates, Bertrand competition on prices embodies competition on these metrics.

\[28\] In each cases considered, regardless of the starting point of strategy, the best response solution from the maximization problem converges to either 0 and 10.
Figure 2.13: Supply and demand under monopolist optimal dynamic pricing solution.
Figure 2.14: Waiting times and request fulfillment rates under monopolist optimal dynamic pricing solutions.
the payoff (platform fare earnings) across combinations of market shares for the two
dispatchers considered, and figures in bold represents the Nash Equilibrium payoffs.

**Table 2.7:** Pure strategy payoffs.

<table>
<thead>
<tr>
<th></th>
<th>case 1</th>
<th></th>
<th>case 2</th>
<th></th>
<th>case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
<td>50%</td>
<td>60%</td>
<td>40%</td>
<td>70%</td>
</tr>
<tr>
<td>s, s</td>
<td>38,188</td>
<td>38,188</td>
<td>47,648</td>
<td>30,190</td>
<td>56,710</td>
</tr>
<tr>
<td>s, ns</td>
<td>39,041</td>
<td>38,157</td>
<td>47,628</td>
<td>30,220</td>
<td>56,706</td>
</tr>
<tr>
<td>ns, s</td>
<td>38,157</td>
<td>39,041</td>
<td>46,694</td>
<td>29,920</td>
<td>54,551</td>
</tr>
<tr>
<td>ns, ns</td>
<td>37,879</td>
<td>37,879</td>
<td>46,688</td>
<td>29,930</td>
<td>54,553</td>
</tr>
</tbody>
</table>

**Table 2.8:** Outcomes of duopoly equilibrium under surge pricing competition.

<table>
<thead>
<tr>
<th></th>
<th>benchmark</th>
<th>case 1</th>
<th></th>
<th>case 2</th>
<th></th>
<th>case 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
<td>50%</td>
<td>60%</td>
<td>40%</td>
<td>70%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>Demand</td>
<td>3,524</td>
<td>3,433</td>
<td>4,148</td>
<td>2,757</td>
<td>4,826</td>
<td>2,151</td>
<td></td>
</tr>
<tr>
<td>Matches</td>
<td>3,158</td>
<td>3,167</td>
<td>3,889</td>
<td>2,519</td>
<td>4,611</td>
<td>1,896</td>
<td></td>
</tr>
<tr>
<td>Pickup prob</td>
<td>0.8435</td>
<td>0.8477</td>
<td>0.9043</td>
<td>0.8234</td>
<td>0.9343</td>
<td>0.8221</td>
<td></td>
</tr>
<tr>
<td>Fulfillment rate</td>
<td>0.8963</td>
<td>0.9224</td>
<td>0.9376</td>
<td>0.9135</td>
<td>0.9555</td>
<td>0.8814</td>
<td></td>
</tr>
<tr>
<td>Mean price surge</td>
<td>—</td>
<td>1.1530</td>
<td>1.1549</td>
<td>—</td>
<td>1.1544</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$19,700</td>
<td>$19,706</td>
<td>$24,645</td>
<td>$14,095</td>
<td>$28,245</td>
<td>$9,270</td>
<td></td>
</tr>
<tr>
<td>Platform surplus</td>
<td>$37,879</td>
<td>$38,188</td>
<td>$47,628</td>
<td>$30,220</td>
<td>$56,706</td>
<td>$22,763</td>
<td></td>
</tr>
<tr>
<td>Ψ</td>
<td>46.88</td>
<td>0.50</td>
<td>9.09</td>
<td>42.35</td>
<td>7.1164</td>
<td>24.49</td>
<td></td>
</tr>
<tr>
<td>Total consumer surplus</td>
<td>$39,400</td>
<td>$39,412</td>
<td>$38,740</td>
<td>$37,515</td>
<td>$37,515</td>
<td>$37,515</td>
<td></td>
</tr>
<tr>
<td>Industry earnings</td>
<td>$75,758</td>
<td>$76,376</td>
<td>$77,848</td>
<td>$79,469</td>
<td>$79,469</td>
<td>$79,469</td>
<td></td>
</tr>
<tr>
<td>Social surplus</td>
<td>$115,158</td>
<td>$115,473</td>
<td>$116,588</td>
<td>$116,984</td>
<td>$116,984</td>
<td>$116,984</td>
<td></td>
</tr>
</tbody>
</table>

From table 2.7, in each case the firm with the dominant market share always
chooses to adopt surge pricing, and the gain for the dominant firm from switching to
surge pricing is higher than that for the dominated firm. When the dominant firm
has a higher share of drivers, the difference in profits across strategies for the other
firm becomes insignificant ($9 when the firm has 30% share), as the price surges will
not be triggered often under the disadvantaged platform. Intuitively, as shown in
Fig. 2.16, a firm that is disadvantaged and has insufficient access to drivers cannot

112
provide a competitive request fulfillment rate and competitive waiting times across all locations to generate demand that would induce drivers’ entry. In consequence, decentralized drivers under the disadvantaged platform migrate to Midtown Center (Fig. 2.15), the location with the highest passenger arrival rate (see Fig. 2.8). In all cases, the number of cabs under each network is insufficient to serve all demand at any period. For instance, even with a 70% share of drivers, the dominant network still cannot fulfill all the requests at 17:40 (panel c in Fig. 2.15), but under the same surge-pricing policy a monopolist dispatcher can guarantee sufficient supply at all time periods (Fig. 2.12). On the supply side, industry earnings increase as the market becomes more consolidated and gains for the dominated firm outweigh the losses of the other firm. On the other hand, total consumer surplus decreases as the loss of having a thin market is not compensated by the fact that the other market thickens. As shown in Fig. 2.16, the thinner market exhibits extremely high waiting times and low fulfillment rates. In any case, the industry earnings and total consumer surpluses are lower than in the monopolistic case ($82,143 and $43,245 respectively).

Spatial Pricing

I first considered the case of symmetric firms. The selection rule of equilibria imposes the restriction that two dispatchers that each has 50% share of drivers have identical demand and spatial cab distributions, which follows that the maximization problems for $i = 1, 2$ are identical given $\nu^{-i}$. Because equilibrium prices are fixed point of the composite best response functions, $\nu^{1*} = \nu^{2*}$. Using this corollary I solved for the spatial pricing equilibrium as the solution to the problem $\text{argmax}_\nu \Pi(\nu, \nu)$. For cases of asymmetric firms, I computed equilibria $(\nu^{1*}, \nu^{2*})$ in the spatial pricing space using an iterative approach. Firms engage in an alternating move game in each iteration until each of their strategies converges.
Figure 2.15: Supply and demand under surge pricing duopoly competition.
Figure 2.16: Fulfillment rate and waiting time under surge pricing duopoly competition.
Table 2.9: Optimal spatial pricing multipliers.

<table>
<thead>
<tr>
<th></th>
<th>case 1</th>
<th>case 2</th>
<th>case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
<td>50%</td>
<td>60%</td>
</tr>
<tr>
<td>Midtown Center</td>
<td>1.0943</td>
<td>1.0943</td>
<td>1.0391</td>
</tr>
<tr>
<td>Midtown East</td>
<td>1.0571</td>
<td>1.0571</td>
<td>1.1290</td>
</tr>
<tr>
<td>Midtown West</td>
<td>0.9434</td>
<td>0.9434</td>
<td>1.1000</td>
</tr>
</tbody>
</table>

Table 2.10: Outcomes of duopoly equilibrium under spatial pricing competition.

<table>
<thead>
<tr>
<th></th>
<th>benchmark</th>
<th>case 1</th>
<th>case 2</th>
<th>case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
<td>50%</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Demand</td>
<td>3,524</td>
<td>3,385</td>
<td>3,994</td>
<td>2,771</td>
</tr>
<tr>
<td>Matches</td>
<td>3,158</td>
<td>3,033</td>
<td>3,776</td>
<td>2,507</td>
</tr>
<tr>
<td>Pickup prob</td>
<td>0.8435</td>
<td>0.7903</td>
<td>0.8494</td>
<td>0.8272</td>
</tr>
<tr>
<td>Fulfillment rate</td>
<td>0.8963</td>
<td>0.8959</td>
<td>0.9455</td>
<td>0.9046</td>
</tr>
<tr>
<td>Waiting time</td>
<td>4.377</td>
<td>3.4365</td>
<td>3.735</td>
<td>3.8449</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$19,700</td>
<td>$18,291</td>
<td>$22,574</td>
<td>$11,365</td>
</tr>
<tr>
<td>Platform earnings</td>
<td>$37,879</td>
<td>$39,089</td>
<td>$48,872</td>
<td>$31,417</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>46.88</td>
<td>98.2997</td>
<td>68.6992</td>
<td>52.3323</td>
</tr>
<tr>
<td>Total consumer surplus</td>
<td>$39,400</td>
<td>$36,582</td>
<td>$33,939</td>
<td>$34,326</td>
</tr>
<tr>
<td>Industry earnings</td>
<td>$75,758</td>
<td>$78,178</td>
<td>$80,289</td>
<td>$80,727</td>
</tr>
<tr>
<td>Social surplus</td>
<td>$115,158</td>
<td>$114,760</td>
<td>$114,228</td>
<td>$115,053</td>
</tr>
</tbody>
</table>

The nature of competition under regional-pricing schedules is slightly different than under surge pricing. Under surge pricing, surge prices are not triggered, because under-equilibrium passengers adjust their choices to avoid price surges. Thus, outcomes of competition under surge-pricing rules are explained mostly by waiting times and fulfillment rates, which are determined by the number of cabs available within the region. Therefore, it can be seen in Fig. 2.15 that the platform with fewer cabs loses demand in all locations except in Midtown Center, where drivers concentrate to create a thick market. Under regional pricing, firms also set prices to induce the desired spatial distributions of supply and demand. This allows a smaller firm to offset its disadvantage in network size by setting lower prices. In particular, for the
case of platforms 1 and 2 owning 60% and 40% of the drivers, platform 2 sets lower fare rates than platform 1 across all locations, and it specifically sets a low fare rate in Midtown East to encourage drivers to enter Midtown West, where the distribution of drivers under platform 1 is sparse. This pricing-strategic interaction results in a zonal duopoly equilibrium outcome in which platform 1 is dominant in Midtown East and platform 2 is dominant in Midtown West, while both firms have similar market shares in Midtown Center. When the market concentration becomes higher (70%, 30%), zonal equilibrium allocation is not feasible for the reduced number of drivers for platform 2: the resulting equilibrium outcome is similar to that of the surge-pricing competition case in which platform 2 actively competes only in Midtown Center. As the dominant firm gets a greater market share, the gains in its earnings outweigh loss to the inferior firm, since we see that industry earnings increase as market concentration increases. Total consumer surpluses are lower in concentrated-market cases than in the symmetric-duopoly case. Nevertheless, they are all lower than in the benchmark case for symmetric platforms and flat pricing.

In a comparison of regional pricing and surge pricing under duopoly competition, it is apparent that surge pricing improves consumer and social surpluses as drivers can respond to locational and also intertemporal variation of demand and better serve passengers. Matching friction under surge pricing is lower than under spatial pricing. In a comparison of Table 2.8 and Table 2.10, surge pricing provides a lower Ψ, implying less misallocation. On the supply side, surge pricing allows a dominant platform to press its advantage in driver network size and gain higher earnings, as the equilibrium response of passengers’ fare rates remains similar across platforms. Regional pricing improves a platform’s earnings when its network size is inferior but still able to support a zonal equilibrium.
Figure 2.17: Supply and demand under spatial pricing duopoly competition.
Figure 2.18: Fulfillment rate and waiting time under spatial pricing duopoly competition.
2.7 Conclusion

This chapter studied the welfare implication of implementing dynamic pricing in the e-hail taxicab market. I presented a structural model featuring a dynamic search and match for drivers, a rational transportation choice process for riders, and an allocation problem for dispatch platforms. Using real-time pricing and waiting times from Uber and estimates of the taxicab market obtained from chapter 1, I calibrated this model on an empirical setup in midtown Manhattan and developed an algorithm to solve for competitive equilibria given a pricing policy, allowing for counterfactual simulations of policy experiments. Finally, I conducted experiments of regional pricing and surge pricing under monopolistic and duopolistic competition in the e-hail taxicab market.

I found network externalities in e-hail dispatch platforms. Platforms that had a larger number of drivers offered thicker markets across locations, offering higher request fulfillment rates. Industry earnings and consumer surpluses were both highest when the e-hail market was operated by a monopolist under any pricing scheme. I also found that surge pricing produced higher social surplus in any market structure considered, while a regional pricing scheme was more profitable for platforms. When strategic interactions between duopoly platforms were accounted for, surge-pricing competition allowed the dominant platform to take advantage of its larger networks to gain a higher market share, and regional-pricing competition made room for zonal equilibrium that helped an inferior platform.

There are a number of topics that the model presented in this chapter can be used to address in future work. With better data from ride sharing platforms such as Uber and Lyft, and e-hail taxi dispatchers, it may be possible to develop more detailed models that predict how drivers react to short run demand and supply shocks that
are “off equilibrium” and require updating of beliefs, to see how more adaptive behavior by drivers affect equilibrium outcomes and welfare of drivers and passengers.
Appendix A

Simulated Generalized Method of Moments Estimates
Table A.1: Simulated generalized method of moments estimates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Point Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_I$</td>
<td>Cobb Douglas coefficient for cabs in Midtown</td>
<td>0.5750</td>
</tr>
<tr>
<td>$\alpha_{II}$</td>
<td>Cobb Douglas coefficient for cabs in Downtown</td>
<td>0.6908</td>
</tr>
<tr>
<td>$\alpha_{III}$</td>
<td>Cobb Douglas coefficient for cabs in Uptown</td>
<td>0.5970</td>
</tr>
<tr>
<td>$\alpha_{IV}$</td>
<td>Cobb Douglas coefficient for cabs in other Boroughs</td>
<td>0.4980</td>
</tr>
<tr>
<td>$\alpha_V$</td>
<td>Cobb Douglas coefficient for cabs in the Airports</td>
<td>0.5040</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Location fixed effect for Bloomingdale</td>
<td>0.8018</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Location fixed effect for Upper West Side</td>
<td>1.5283</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Location fixed effect for Lincoln Square</td>
<td>1.4410</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>Location fixed effect for Clinton</td>
<td>1.5994</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>Location fixed effect for Chelsea</td>
<td>1.8344</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>Location fixed effect for Central Park</td>
<td>0.4594</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>Location fixed effect for Times Square</td>
<td>2.5305</td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>Location fixed effect for Penn Station</td>
<td>1.4949</td>
</tr>
<tr>
<td>$\beta_9$</td>
<td>Location fixed effect for Flatiron</td>
<td>1.7883</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>Location fixed effect for Harlem</td>
<td>0.4035</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>Location fixed effect for Yorkville</td>
<td>2.1904</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>Location fixed effect for Lenox Hill</td>
<td>1.9929</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>Location fixed effect for UN</td>
<td>2.0235</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>Location fixed effect for Murray Hill</td>
<td>1.6450</td>
</tr>
<tr>
<td>$\beta_{15}$</td>
<td>Location fixed effect for Gramercy</td>
<td>1.6017</td>
</tr>
<tr>
<td>$\beta_{16}$</td>
<td>Location fixed effect for Greenwich Village</td>
<td>0.9761</td>
</tr>
<tr>
<td>$\beta_{17}$</td>
<td>Location fixed effect for Alphabet City</td>
<td>0.9720</td>
</tr>
<tr>
<td>$\beta_{18}$</td>
<td>Location fixed effect for Lower Manhattan</td>
<td>2.5596</td>
</tr>
<tr>
<td>$\beta_{19}$</td>
<td>Location fixed effect for other Boroughs</td>
<td>5.6810</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>Location fixed effect for JFK and LGA</td>
<td>5.6084</td>
</tr>
<tr>
<td>$\gamma_i^1$</td>
<td>Night time effect for Midtown</td>
<td>0.6652</td>
</tr>
<tr>
<td>$\gamma_{II}^1$</td>
<td>Night time effect for Downtown</td>
<td>0.8607</td>
</tr>
<tr>
<td>$\gamma_{III}^1$</td>
<td>Night time effect for Uptown</td>
<td>0.4010</td>
</tr>
<tr>
<td>$\gamma_{IV}^1$</td>
<td>Night time effect for other Boroughs</td>
<td>1.0322</td>
</tr>
<tr>
<td>$\gamma_i^2$</td>
<td>Night time effect for JFK and LGA</td>
<td>0.7302</td>
</tr>
<tr>
<td>$\gamma_{II}^2$</td>
<td>Morning peak effect for Midtown</td>
<td>1.1334</td>
</tr>
<tr>
<td>$\gamma_{III}^2$</td>
<td>Morning peak effect for Downtown</td>
<td>1.0834</td>
</tr>
<tr>
<td>$\gamma_{IV}^2$</td>
<td>Morning peak effect for Uptown</td>
<td>1.6873</td>
</tr>
<tr>
<td>$\gamma_i^3$</td>
<td>Morning peak effect for other Boroughs</td>
<td>3.3621</td>
</tr>
<tr>
<td>$\gamma_{II}^3$</td>
<td>Morning peak effect for JFK and LGA</td>
<td>1.1492</td>
</tr>
<tr>
<td>$\gamma_{III}^3$</td>
<td>Evening peak effect for Midtown</td>
<td>1.1707</td>
</tr>
<tr>
<td>$\gamma_{IV}^3$</td>
<td>Evening peak effect for Downtown</td>
<td>1.0047</td>
</tr>
<tr>
<td>$\gamma_i^4$</td>
<td>Evening peak effect for Uptown</td>
<td>1.1843</td>
</tr>
<tr>
<td>$\gamma_{II}^4$</td>
<td>Evening peak effect for other Boroughs</td>
<td>1.2032</td>
</tr>
<tr>
<td>$\gamma_{III}^4$</td>
<td>Evening peak effect for JFK and LGA</td>
<td>1.3159</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Scale parameter for unobserved utility for street hail drivers</td>
<td>3.8849</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Network parameter for e-hail matching function</td>
<td>5.3689</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>Scale parameter for unobserved utility for e-hail drivers</td>
<td>1.0988</td>
</tr>
<tr>
<td>$G_e$</td>
<td>Percentage of e-hail driver among green cabs</td>
<td>0.3522</td>
</tr>
</tbody>
</table>
Appendix B

An Experiment on Carpooling

On top of acting as a dispatcher, e-hail programs are able act as a centralized matching platform providing extra rides for passengers who are willing to carpool with other passengers on their way. I conduct a counterfactual experiment assuming the e-hail demand grows by 30% from the data equilibrium and 25% of the cabs are equipped e-hail programs. Suppose \( w \) of e-hail passengers are willing to carpool and \( \psi(s) = w^2 (1 - (1 - \nu)^{\kappa}) \sum_{s'} Q(s, s') \sum_{s'' \in \Pi(s')} Q(s, s'') \), the matching function for \( C^e \) cabs and \( D^e \) e-hail passengers is given by:

\[
m^e = (1 + \psi)(1 - (1 - \nu)^{\kappa}) \min\left\{ \frac{D^e}{1 + \psi}, C^e \right\}
\] (B.1)

The idea behind the matching function is that conditional on picking up an e-hail passenger at state \( s \) who is going to state \( s' \) and who is willing to carpool with another passengers is probability \( w(1 - (1 - \nu)^{\kappa}) \). The next request originating at state \( s \) has a probability of \( w \sum_{s' \in \Pi(s)} Q(s, s') \) for a passenger that is eligible for carpool: he is also going to a set of eligible carpool states \( \Pi(s) \) (adjacent zones) from \( s \) and is willing to carpool. The driver is matched with the second passenger if he is nearby, with probability \( 1 - (1 - \nu)^{\kappa} \). Here I implicitly assumed that a carpool ride only takes at most 2 requests, and that carpooling passengers have the same originating zones.

At the same number of e-hail requests, carpooling decreases the market thickness: passengers can possibly matched with drivers that are already matched with one passenger, so a vacant cab faces more competition. This can be seen that now the effective
market thickness parameter is $\min \{\frac{D^e_e}{1+\psi}, C^e\} < \min \{D^e, C^e\}$. On the other hand, carpooling reduces search frictions, as it is seen from the e-hail carpool matching function that it is possible to clear the demand when there is a shortage of e-hail cabs. Under the lower level demand for e-hail rides observed in the data, demand-supply ratio is not binding and carpooling does not provide significant improvement. When e-hail demand substitutes higher proportions of street hail demand, carpooling generates significant gains in matching efficiency. Because I do not have access to carpooling travel time and passenger usage data in any form, it would not be feasible to model endogenous response from passengers. Instead, I consider exogenous percentages of passengers who are willing to carpool.

Table B.1: Percentage increase in matches for each given % of street hail demand substituted.

<table>
<thead>
<tr>
<th>% of passengers willing to carpool</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>1.63%</td>
<td>3.05%</td>
<td>3.52%</td>
<td>4.34%</td>
<td>4.95%</td>
</tr>
<tr>
<td>40%</td>
<td>2.59%</td>
<td>4.84%</td>
<td>7.85%</td>
<td>8.86%</td>
<td>10.85%</td>
</tr>
<tr>
<td>50%</td>
<td>4.55%</td>
<td>7.52%</td>
<td>9.93%</td>
<td>11.71%</td>
<td>15.33%</td>
</tr>
</tbody>
</table>
A Discussion on the Parametric Forms of Matching Functions

This section of the appendix discusses the parametric forms on the street-hail and e-hail matching functions. Parametric assumptions are made to decrease the dimensionality of competitive equilibria, making estimation computationally more feasible.

C.1 Street-Hail Matching Function

Assume that the encounter between a searching cab and a waiting passenger follows a Poisson($\lambda$) arrival process, the probability that the searching time is smaller than $x$ is $1 - \exp(-\lambda x)$. Because the event of a driver matched within timespan $x$ is equivalent to the event that his searching time is less than $x$, therefore: $P($pickup within $x) = 1 - \exp(-\lambda x)$. The assumption of Poisson arrival rate is consistent with the findings in the 2013 panel TLC trip record, in which drivers’ searching times are exponential distributed as shown in the graph below.
Recall that in aggregate, the street hail matching process with $C$ searching vacant cabs and $D$ waiting passengers follows a Cobb-Douglas function: $m(C,D) = AC^aD^b$. I chose a Cobb-Douglas functional form because simulations of moving searching cabs and waiting passengers in a rectangular lettuce with geographical passenger arrival rate variation generate data is closely fitted by a Cobb-Douglas form matching function, as discussed in appendix D.

$$P(z,t) = 1 - \exp(-\lambda_z^t x_z^t) = (C_z^t)^{\alpha_z-1}\beta_z \gamma_z,t$$

$$\lambda_z^t = -\frac{1}{x_z^t} \log(1 - (C_z^t)^{\alpha_z-1}\beta_z \gamma_z,t) := F(C_z^t; \theta)$$

Figure C.1: Empirical density of cab drivers in 2013 and reference exponential density function.

Notice that since $\alpha_z$ is less than 1, increasing the number of searching cabs at $(z,t)$, $C_z^t$ increases the number of matches produced, but decreases the likelihood of each individual driver of picking up a passenger. Recall that the variable $x_z^t$ is defined
as the time a cab must stay in its originating zone. It is also the portion of the period $t$ during which the number of searching cabs are relatively fixed.

The parameters $\alpha_z$ is interpreted as the efficiency of matching at zone $z$, geographical variation in matching efficiency takes into account road structures and existence of taxi stands. $\beta, \gamma$ capture the demand of $(z, t)$ within the $x$ minutes: $\beta_z$ can be interpreted as the base demand of the zone $z$, while $\gamma^t_z$ is the "surge factor" of demand, representing the spatial response to high and low peaks at different periods of the day. $\beta_z$ is specific to each zone $z$, and zones in the same district\textsuperscript{1} shares the same matching efficiency parameter $\alpha_z$ and demand surge multiplier $\gamma^t_z$. The district partition of the map space is shown in the figure below. On the time dimension, $\gamma^t_z$ vary in three phases: midnight (12 a.m. – 7a.m.), morning peak (7 a.m. – 10 a.m.) and evening peak (4 p.m. – 8 p.m.), and the multiplier for the remaining time periods is normalized as 1.

The parametric form of the matching probabilities describes how spatial distribution of cab induced by their searching decisions and the pickup probability across the map are determined by each other. When drivers’ decision rules induce spatial distribution of searching cabs given by $C^t_z$, the function $F(C^t_z; \theta)$ where $\theta = (\{\alpha_z\}_z, \{\beta_z\}_z, \{\gamma^t_z\}_{z,t})$ is the parameter vector will determine the individual matching probability parameter $\lambda$. When drivers are able to observe and gather information about changes in other cabs’ behavior, they can adjust their search decisions accordingly as given by the function $F$.

\textsuperscript{1} Zones are grouped into larger districts that they belong to, including Midtown, Downtown, Uptown, other Boroughs and the Airports.
C.2 E-hail Matching Function

The matching function for e-hail cabs is derived from an e-hail matching algorithm as follows: for each e-hail request, the dispatch platform generates a list of $\kappa$ e-hail partner cabs in the same zone. The request is matched with the cab ranked top nearby the requested location, with probability $\nu(z,t)$. This is consistent with the TLC regulation that e-hail matches must be made within 0.5 miles in downtown, 1.5 miles elsewhere and the fact that zones in the model are larger than 1.5 square miles. Therefore, probability that a request gets matched will be the probability that at least one cab in the $\kappa$ e-hail cabs in the same zone is within matching distance. It
follows that the probability that the request gets matched is $1 - (1 - \nu(z, t))^\kappa$ and the aggregate matching function is:

\[ m^e(C^e, D^e) = \min\{C^e, D^e\}(1 - (1 - \nu(z, t))^\kappa) \]  \hspace{1cm} (C.3)

And from the perspective of an e-hail driver the pickup probability is:

\[ P^e(C^e, D^e) = \min\left\{1, \frac{D^e}{C^e}(1 - (1 - \nu(z, t))^\kappa)\right\} := \min\left\{1, \lambda^e(1 - (1 - \nu(z, t))^\kappa)\right\} \]  \hspace{1cm} (C.4)

I refer to the parameter $\kappa$ as the network strength of the average e-hail taxi dispatch platform. When the network strength diverges to $\infty$, the e-hail matching function approaches the frictionless matching function $\min\{C^e, D^e\}$. $\kappa$ would depend on the number of cab drivers the average dispatcher platform has access to. For example, a high number of firms might cause a low value of $\kappa$ because the number of cars reached by each request on average is decreased, when each e-hail driver only partners with one e-hail dispatcher. For instance, Uber has a much larger driver network than the average e-hail cab platform, as shown in the figures below. When the meeting probability $\nu(z, t)$ are calibrated, I use the parameter vector $\lambda^e$ in the fixed point algorithm to approximate the e-hail drivers’ beliefs of getting requests. $\lambda^e = \frac{D^e}{C^e}$ corresponds to the inverse market tightness at $(z, t)$. Analogous to equilibrium with only street hail cabs, market tightness is a function of distribution of e-hail cabs and therefore defines a spatial matching game.
Figure C.3: E-hail cab network ($\kappa = 5.36$).

Figure C.4: Uber network ($\kappa >> 10$).
Appendix D

A Discussion on Equilibrium Convergence

Recall that existence of equilibrium is equivalent to the existence of a fixed point for $\Gamma$, the operator mapping drivers’ pick up probabilities to actualized probabilities induced by the vacant cab distribution under their optimized policy. This appendix section provides a more analytical representation for $\Gamma$ and shows that an equilibrium exists, and demonstrates in a simplified game in which $\Gamma$ can be expressed analytically successive belief iteration results in a fixed point.

D.1 Continuity of $\Gamma$

Recall that $\Gamma$ maps a belief vector of pickup probabilities $P$ to an actualized pickup probability vector $P_2$. Denote vacant cab distribution under optimal policy with belief $P$ as $C(P)$:

$$P_2 = \frac{m(C(P))}{C(P)} := F(C(P))$$  \hspace{1cm} (D.1)

Since $m$ is continuous as shown in appendix C and $C > 0$ for any all location and time because the numbers of observed matches are non-zero, $F$ is continuous in $C(P)$. Next, it is noted that the vacant cab distribution at $(z,t)$ is the sum of net new entry of cabs in the market (drivers starting their shifts at $(z,t)$, cabs just dropped off passengers at $(z,t)$ and cabs that searched at $z$ from last period but did not get matched. Express this accounting equations as:
\[
C_t^z(P) = \tilde{X}_t^z + \left(1 - \rho(X_t^z)\right) \left\{ \sum_{z',t'} m_{z,t'}^t Q(z, t | z', t') \right. \\
\left. + \sum_{z'} C_t^{z-1}(P) \left[1 - F(C_t^{z-1}(P))\right] \phi(z', t, z; P) \left[1 - F(C_t^{z-1}(P))\right] \right\} \tag{D.2}
\]

Cabs did not matched in first \(x\) minutes in \(z'\) and \(10 - x\) minutes in \(z\)

Because the initial states for cabs in terms of shift starting time and zones are given, \(\tilde{X}_t^z \) and \(X_t^z\) are exogenous. Furthermore, if an initial point \(C_0^z\) along the path of \(\{C_t^z\}\) is given, then dependency on \(P\) of \(C_t^z(P)\) is through \(\phi\). However recall that because unobserved utility terms are Gumbel distributed the optimal policy is a in the logit form of the action-specific utility \(W(z', t, z)\):

\[
\phi(z', t, z; P) = \frac{\exp(W(z', t, z; P))}{\sum_z \exp(W(z', t, z; P))} \tag{D.4}
\]

\(\phi\) is continuous in \(W\) which is a linear function in \(P\). Therefore, \(\Gamma\) is continuous in \(P\) and by the Brouwer fixed point theorem a fixed point exists for \(\Gamma\).

D.2 Successive Belief Iteration in a Static Matching Game

Let there be \(n\) zones \(\{z_i\}_{i=1}^n\). At \(t = 0\) there are \(C_z\) cabs at zone \(z\) at the beginning of the game. Each cab driver decides where he would like to search for passengers, when he is matched with a passenger he receives a payoff of 1. Traveling to the other zone incurs a disutility of \(-c\). The unobserved component of the action specific utilities is given by \(\varepsilon\), EV type 1 distributed. Write the action specific value function as:

\[
V_z(z') = -c I\{z \neq z'\} + P(\text{ride}|z') + \varepsilon_z(z') \tag{D.5}
\]
Naturally, the probability of matching is a function of the actual number of searching cabs, so suppose the law of motion for cab drivers is given by $P(z|z') = \phi_{z'}(z)$:

$$P(\text{ride}|z) = F(\bar{C}_z, \phi) = F\left(\sum_{z'} \phi_{z'}(z)C_{z'}\right)$$  \hspace{1cm} (D.6)

Here $F(C|z) = \frac{m(C|z)}{C}$ where $m(C|z)$ is the matching function for zone $z$. Without loss of generality, $F$ can permit passengers’ endogenous responses, e.g. passengers respond to waiting time variations which are functions of $C$. Following the terminology in the paper, a competitive equilibrium definition is:

**Definition:** A competitive equilibrium in this model is: belief $\tilde{\phi}$, decision rule $\phi$ such that:

- $\phi$ solve the drivers’ optimization problem.
- Drivers’ belief $\tilde{\phi}$ is consistent with the law of motion of vacant cabs induced by optimal strategy $\phi$.

Optimality of equilibrium requires that $\phi$ follows the softmax form:

$$\phi_{z'}(z) = \Phi_{z'}(z, \tilde{\phi}) = \frac{\exp\left(\frac{V_{z'}(z, \tilde{\phi})}{\sigma}\right)}{\sum_z \exp\left(\frac{V_{z'}(z, \tilde{\phi})}{\sigma}\right)}$$  \hspace{1cm} (D.7)

And self-fulfilling implies that at equilibrium:

$$\phi = \tilde{\phi}$$  \hspace{1cm} (D.8)

The goal for this exercise is to show that an equilibrium can be computed by updating the current episode’s driver beliefs $\tilde{\phi}$ with the utility maximizing law of
motion from the previous episode \( \phi \). First the following results from Gao & Pavel (2017) is invoked.

**Proposition:** The softmax function \( \Phi(V, \sigma) = \frac{\exp\left(\frac{V}{\sigma}\right)}{\sum \exp\left(\frac{V'}{\sigma}\right)} \) is \( \frac{1}{\sigma} \)-Lipschitz in \( V \):

\[
\| \phi(V) - \phi(V') \|_2 \leq \frac{1}{\sigma} \| V - V' \|_2
\]

The proof for the above proposition utilizes the Baillon-Haddad theorem and that the log-sum-exponential function is convex and monotone.

Given an initial belief \( \tilde{\phi}_0 \), define \( \Gamma : \Delta^1 \to \Delta^1 \) the "belief updator operator" such that that \( \Gamma(\tilde{\phi}_0) = \Phi(V(\tilde{\phi})) = \tilde{\phi}_1 \) given by the softmax transformation (D.2). In other words, \( \Gamma \) represents the process in which players update their beliefs in the process of playing the repeated matching game in the following way: the players start with a (common) initial belief of the strategies of other players and then in the next episode, update their beliefs with the realized behavior in the previous episode. In machine learning terminology, this corresponds to a reinforcement learning process with learning rate 1.\(^1\) It can be shown that in \( \Gamma \) is continuous in \( \phi \) when \( F \) is continuous, therefore the existence of fixed point is guaranteed by Brouwer’s Fixed Point Theorem.

**Lemma: (Sufficiency of convergence)** \( \Gamma \) is a contraction mapping in \( \phi \) under the following condition:

\[
\frac{1}{\sigma} \| \nabla_{\phi} F(\phi \cdot C) \| < 1 \quad \forall \phi \in \Delta^1
\]

\(^1\)An intuitive conjecture is that since that the (expected) initial cab distribution is set to be fixed and beliefs are common among drivers, so a learning rate of 1 would suffice
**Proof:** The result from Gao & Pavel (2017) states that the softmax function \( \Phi \) is \( \frac{1}{\sigma} \)-Lipschitz on the payoff function space. Therefore,

\[
\| \Gamma(\tilde{\phi}_1) - \Gamma(\tilde{\phi}_0) \|
\]

\[
= \| \Phi(V(\tilde{\phi}_1)) - \Phi(V(\tilde{\phi}_0)) \|
\]

\[
\leq \frac{1}{\sigma} \| V(\tilde{\phi}_1) - V(\tilde{\phi}_0) \|
\]

\[
= \frac{1}{\sigma} \| (F(\hat{\phi}_1 \cdot C) - c_{\lambda}) - (F(\hat{\phi}_0 \cdot C) - c_{\lambda}) \|
\]

\[
\leq \frac{1}{\sigma} \| \nabla_{\phi} F(\hat{\phi} \cdot C) \| \| \hat{\phi}_1 - \hat{\phi}_0 \|
\]  

(D.11)

Where \( \hat{\phi} = \delta \Phi(V(p_1)) + (1 - \delta) \Phi(V(p_2)) \) for some \( \delta \in (0, 1) \). The last inequality invokes mean value theorem for multivariate functions and the Cauchy-Schwartz inequality.

**Corollary:** Under the same condition, the updator operator \( \Gamma^\lambda \) for a parameter \( \lambda \) with \( G(\lambda) = P(\text{ride}) \) also exhibits contractive properties if \( G \) is \( 1 \)-Lipschitz.

**Proof:** Since \( G \) is \( 1 \)-Lipschitz, it suffices to show that the updator operator \( \Gamma^p \) is contractive in \( p \):

\[
\| \Gamma^p(p_1) - \Gamma^p(p_2) \|
\]

\[
= \| F(\Phi(V(p_1)) \cdot C) - F(\Phi(V(p_2)) \cdot C) \|
\]

\[
\leq \| \nabla_{\phi} F(\hat{\phi} \cdot C) \| \| \Phi(V(p_1)) - \Phi(V(p_2)) \|
\]

\[
\leq \| \nabla_{\phi} F(\hat{\phi} \cdot C) \| \sigma \| V(p_1) - V(p_2) \|
\]

\[
= \| \nabla_{\phi} F(\hat{\phi} \cdot C) \| \sigma \| p_1 - p_2 \|
\]  

(D.12)

Where \( \hat{\phi} = \delta \Phi(V(p_1)) + (1 - \delta) \Phi(V(p_2)) \) for some \( \delta \in (0, 1) \). This corollary implies that instead of having to account for behaviors of the representative driver, it is sufficient to have a consistent belief system on the probability of pick up which is in
a lower dimension, or equivalently, a set of parameter that can be mapped through a function to \( p \) that is 1-Lipschitz on the parameter domain, one of such examples would be \( C \), the distribution of cabs which is mapped to the probability space through \( F \).

**Example:** The matching function used in the estimation is \( m(C) = KC^\alpha \), so \( F(C) = KC^{\alpha - 1} \). The logit parameters estimated are \( \hat{\alpha} \) around 0.4 to 0.6 for different areas and \( \hat{\sigma} \) range 4 and the multiplicative constant lies between 1 to 18. The total number of active cabs at any time is bounded between 2,000 to 13,000. Simulation results show that the condition is always satisfied under the range of parameters.

**Visualization:** The fixed point convergence for the full model with respect to the matching parameter \( \lambda \) is shown in the paper. To visualize the fixed points, consider the even simpler version of the static game: all \( C \) cabs come from the same initial zone \( z_0 \) (which they cannot match) and decide to go one of \( z_1 \) and \( z_2 \). Matching probabilities are \( F_1(C) = C^{\alpha_1} \) and \( F_2(C) = C^{\alpha_2} \) where \( \alpha_i \in (-1, 0) \). Payoffs are identical to \( F_i \). Below shows the belief updator \( \Gamma(\phi_1) \) for belief going to \( z_1 \), \( \phi_1 \) when \( (\alpha_1, \alpha_2) = (-0.8, -0.4) \), \( C = 100 \) for \( \sigma \) 0.1 (blue) and 1 (red). One can verify that condition (3) is satisfied for the latter case, it is worth noting that the belief updator manages to contract even when sufficiency (3) is not always satisfied for \( \sigma = 0.1 \).
Note that the results from Gao & Pavel (2017) shows that it suffices to show that the normalized payoff function is also contracting.

D.3 Dynamic Matching Game

Extending from the static matching game above to a dynamic setting, the following Bellman equation represents the dynamic programming problem for a driver at zone $z \in \{1, \ldots, n\}$ and time period $1 \leq t \leq T$:
Without loss of generality, consider the case $T = 2$ and assume that drivers do not return to the map once they are matched with a passenger. Also noting that the unobserved shock is extreme value distributed:

\[
V_z(z', t) = -cI\{z' \neq z\} + P(\text{ride}|z') + \beta(1 - P(\text{ride}|z'))V_{z'}(z, t) + \varepsilon(z') \tag{D.13}
\]

Where $lse(\cdot)$ is the log-sum-exp function. As seen in the static case that the belief updator is contracting for second period $V_z(z', 2|\tilde{\phi})$ if the normalized norm of $\phi$ gradient of $G(\tilde{\phi}) = F(\tilde{\phi}^2 \cdot (1 - F(\tilde{\phi}^1 \cdot C)) \cdot \tilde{\phi}^1 \cdot C)$ is strictly less than 1, i.e. $\frac{\|\nabla_G\|}{\sigma} < 1$. Note that as it is shown that conditions that $V_z(z', 2|\tilde{\phi})$ is contracting, it is sufficient to pin down the sufficient condition for which $V_z(z, 1|\tilde{\phi})$ is also contracting, because $lse$ is contracting, bounded and $V_z(z', 1|\tilde{\phi})$ is the sum of terms which are either Lipschitz in $\tilde{\phi}$ or product of bounded Lipschitz functions. When the composite Lipschitz constant for $V_z(z', 1|\tilde{\phi})$ is sufficiently small it can be easily shown that the belief updator is contracting.
APPENDIX E

DETAILS ON DEMAND DATA IMPUTATION

E.1 Estimation

The main challenge for estimating the nested logit model of transportation choice is that destinations of Uber and MTA passengers are not observed in the data, which also causes the exact prices paid and total travel times to be missing as well. Using the notations in Rubin (1987), let $Y^i = (choice^i, origin^i, destination^i, \{(p, Et)^i_j\})$ denote the vector of observation of passenger $i$ and $Y_{mis}^i, Y_{obs}^i$ be the observation vectors of $Y$ without and with missing observations. Further denote $R^i_j$ the indicator that the $j$-th field of $Y^i$ is observed. Since the observation of alternative prices and waiting times are conditional on whether the passenger chosen to ride taxi, the likelihood function of $f(R^i_j)$ for $j > 2$ is $I\{\text{choice}^i = \text{taxi}\}$. The likelihood function of data with missing observations in terms of model parameter $\theta$ is:

$$f(Y_{obs}, R|\theta) = \int f(Y_{obs}, Y_{mis}|\theta)f(R|Y_{obs}, Y_{mis}, \theta)dY_{mis}$$
$$= \int f(Y_{obs}, Y_{mis}|\theta)f(R|Y_{obs})dY_{mis}$$
$$= \int f(Y_{obs}, Y_{mis}|\theta)dY_{mis} \quad (E.1)$$

Because events of missing of price and travel time data is the same as the event that passengers chose to ride Uber or metro, the missing pattern is deterministic conditional on the observed choices. According to the above nested discrete choice model, the likelihood function of the data is:
\[
\int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta)dY_{\text{mis}} = \int \prod_{i,j=\text{choice}^i} \frac{\exp(a_{k(j)} X^i) \exp(IV_{k(j)})}{\sum_{k' \in K} \exp(a_{k'} X^i) \exp(IV_{k'})} \times \exp\left(\frac{(bp^i_j + cEt^i_j)/\xi_{k(j)}}{\sum_{j' \in k(j)} \exp\left(\frac{(bp^i_{j'} + cEt^i_{j'})/\xi_{k(j')}}{\xi_{k(j')}}\right)}\right)dY_{\text{mis}} \quad (E.2)
\]

Under the aforementioned assumptions on passengers flow and waiting times, the posterior distributions of the destinations, prices and travel times are identified and values of missing data can be imputed. The likelihood function at a given parameter value \(\theta\) can be approximated by evaluating the integrand over a large sample of imputed data. It is noted that maximum likelihood estimation can be very computationally expensive, as the evaluation of integration is subject to a multivariate distribution for each field with missing data. In light of this, I propose the following bootstrapped multiple imputation estimation:

**Algorithm 5** Multiple imputation.

1: From the observed dataset consisting of time, location and choice for riders, sample \(M\) subsamples of \(N\) riders.
2: for \(m = 1, \ldots, M\) do
3: Draw destinations for riders under the posterior distribution conditional on their choices, time and locations.
4: Impute the corresponding taxi, Uber, MTA fares and travel times.
5: Estimate demand parameters \(\mu_m\) for dataset \(m\), matching the market shares of taxi and Uber for each \((\text{zone}, \text{hour})\) pair of sample \(m\).
6: end for
7: Combine and summarize on estimates \((\hat{\mu}_1, \ldots, \hat{\mu}_M)\) to obtain point estimates and confidence intervals.

Where in the last step point estimates are obtained from the mean \(\frac{1}{M} \sum_{m=1}^{M} \hat{\mu}_m\) and the standard errors of imputed method of moments estimator in the last step follows Kim & Yang (2016).
E.2 Origin-Destinations Imputation

The metropolitan transport authority conducted an public transit origin destination survey on 3,275,607 adults in 2008, among which 1,598,496 are subway riders. Of the 49,167 weekday trips the MTA provided, departure times, census tract numbers of origins and destinations for the passengers were reported, identifying the law of motions of metro riders. Due to the limited number of available data it is not feasible to partition the distribution for every hour, instead, I partition the time space of a day to morning peak (6 a.m. to 11 a.m.), evening peak (4 p.m. to 7 p.m.) and no peak. For each Uber rider observed originating from zone $z$ at hour $h$, the destination was imputed using the conditional probability observed in the 2013 taxicab trip record dataset.

E.3 Fare Imputation

I applied the obtained law of motions to the sample of full-fare metro riders to construct the estimation sample. Dropping metro riders using 7- and 30- day passes prevents unobserved fare heterogeneity, as their metro fare were paid upfront as a sunk cost and the joint distribution of the destination and pass users are not observed. The full fare of metro subway is fixed at $2.75.

Using assumption (3), the driving distance and duration for each (origin, destination, hour) tuple is drawn from a normal distribution whose mean and variance parameters are calibrated from the 2016 taxicab data set, and the fare is computed using the fare formula, including any surge multiplier applied. For taxicabs fare imputation, since the fare of taxi trips are depends on the speed of traffic and how often the driver stopped, instead of computing the fare for taxi rides as a function of distance
and duration, the imputed fares are drawn from the empirical conditional distribution of taxicab fare. The price for taxi rides include all payments made to the driver, including tips and tolls.

E.4 Travel Time Imputation

The travel time of metro is the sum of walking time from origin location to metro station, waiting time and transit time. The zone-to-zone travel time excluding metro waiting time was collected from Google Maps database through its API, where the exact geo-coordinates of origin and destination zones are represented by the default location inputting the name of the zone. Assuming uniform arrival rates for passengers, the expected waiting times are half of the inter arrival times for trains, which are obtained from the MTA schedules.

For ride hailing services, travel time is the sum of waiting time and trip duration. Conditional on origin, destination and time of pick up, the trip duration for taxi and Uber trips are drawn from the empirical distribution of trip duration from the taxi trip records. The real time Uber waiting time is directly collected from the Uber API, whereas the distribution of taxi waiting time is estimated from within zone motion of cabs, as described in the appendix.

Table E.1: Statistics of trip characteristics in imputed data.

<table>
<thead>
<tr>
<th></th>
<th>taxi</th>
<th>uber</th>
<th>metro</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance</td>
<td>2.36</td>
<td>2.39</td>
<td>4.71</td>
</tr>
<tr>
<td>duration</td>
<td>21.46</td>
<td>7.922</td>
<td>33.61</td>
</tr>
<tr>
<td>fare</td>
<td>$14.67</td>
<td>$9.89</td>
<td>$2.75</td>
</tr>
<tr>
<td>market share</td>
<td>29.04%</td>
<td>11.06%</td>
<td>59.90%</td>
</tr>
</tbody>
</table>
Appendix F

Details on Waiting Time Simulation

F.1 Environment

To obtain estimates of taxi waiting time, a microfounded simulation environment based on the set up of Frachette et al. (2017) is adapted to approximate the waiting time for taxi passengers. Note that the street block structure of Midtown, Uptown and (North) downtown are rectangular, the distance between two blocks along the North-South and East-West directions are 1/20 and 1/5 miles respectively. Each zone in the model is nearly rectangular and therefore I express each zone as a matrix, where the distance between two points is 1/20 miles. While the model accounts for taxi drivers’ behavior across the New York City, the goal is to simulate passenger arrivals and motions of cabs within each zone to recover passengers’ waiting times.

I acknowledge that each zone has different geographical and road structures and I attempt to replicate them. For instance, roads are one-way and cabs need to follow the direction of each road and street on the matrix that corresponds to the street in real life. At each discrete time period that correspond to the time needed for cab to move one grid point (\( \frac{2}{v} \) minutes where \( v \) is the speed of traffic), waiting passengers arrive at each point following a Poisson(\( \lambda \)) process. It is noted that there are heterogeneities of arrival rates in each zone, as reflected by the spatial pick up data. For example, the map representation of pick up in Upper West Side shows that pick ups happened more frequently along avenues. And the data of Murray Hill shows that pick ups tend
to take place on the west side of the zone. The features are incorporated in the model by giving the corresponding points higher Poisson arrival parameters.

![Figure F.1: Upper West Side.](image1)

![Figure F.2: Murray Hill.](image2)

**F.2 THE HEAT MAP RULE - INFERRING TAXI WAITING TIME IN MANHATTAN**

Acknowledging heterogeneous passenger arrival rates within a zone, the following "heat map" rule for drivers’ movements. Denote $\Lambda(i, j)$ the passenger arrival rate at a point $(i, j)$ on the map matrix and the driver of a vacant cab is at junction $(a, b)$ decides if he is searching in North-South ($y$) or East-West ($x$) direction. Further define $d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$ the "Manhattan distance" between two points $(x_1, y_1)$ and $(x_2, y_2)$ and the $D(a, b, F, z) = \{(i, j) : d((x, y), (a, b)) \leq F\}$ the set of points within $F$ steps in direction $z$. I define the "heat" in the $z$ direction be:

$$F_z(a, b, F) = \sum_{i, j \in D(a, b, F, z)} \frac{\Lambda(i, j)}{d((a, b), (i, j))}$$  \hspace{1cm} (F.1)

Since roads are one way in the map space, the driver faces at most two heats, along the North-South and East-West directions at road intersections. Suppose the driver is at a road intersection that he chooses to go either North or East, the probability of driver choosing to go on the N (resp. E) direction is set to be:

$$P_N = \frac{F_N}{F_N + F_E}$$  \hspace{1cm} (F.2)
I refer the law of motions for vacant cab drivers defined by equations F.2 and F.2 the "heat map" rule.

F.2.1 Simulation

The heat map rule applies to zones that exhibit the rectangular, uniform street block structures, including Bloomingdale, Upper West Side, Lincoln Sq, Clinton, Chelsea, Midtown, Penn Station, Union Square, E Harlem, Upper East Side, Gramercy and East village. The parameter for number of steps ahead considered, $F$ is chosen to be 4, so drivers are planning one block ahead. The geographical characteristics of dimensions, road locations and arrival heterogeneities are translated into a matrix and simulations were run in the following steps:

**Algorithm 6 Heatmap Simulation for a (zone, time) pair.**

1: From the model solution obtain the equilibrium number of cabs and the probability of staying in the zone.
2: Generate heat map rule probabilities and Poisson arrival rates at each point on the map.
3: for $t = 1,...,T$ do
4: Passengers arrive at each point $(i,j)$.
5: Simulate cab movements across the map matrix according to the heat map rule.
6: Match cab and passengers when a cab and a passengers are at the same matrix point. If there are multiple passengers at a point, the passenger who has waited the longest gets matched for each cab.
7: A passenger leaves the map if she has waited for more than 15 minutes.
8: When a cab has stayed on the map for more than $x(z,t)$ minutes, it leaves or stays according to the equilibrium policy.
9: end for
10: Obtain the distribution of waiting time over $N$ continuous iterations of the above loop.

Recall that in the model, each vacant cab originating from a zone $z$ will stay in the zone for $x(z,t)$ minutes, regardless of the drivers’ search decisions. To identify
the waiting time of taxi riders in the data, I estimate the total arrival rates in a zone that matches the number of pick ups that appeared in the trip record data. Under the solution of arrival rates, the average waiting time is illustrated in the following figure:

Figure F.3: Simulated waiting times for taxicabs (solid) vs. Uber (dotted) in Manhattan.

The yellow bold lines correspond to the simulated average passenger waiting time. The waiting time pattern for taxi is consistent with the overall supply of cabs relative to the demand. At midnight where the cab supply is minimal, passengers have long waiting time. Another peak of waiting time appears around 5:00 p.m., where the end of day shifts take place but commuter demand surges. Where the black dotted line corresponds to the Uber waiting time. On average, taxi waiting time is shorter than Uber, but is more sensitive to supply shock during the evening peaks. It is a standard wisdom among New York City residents that in Manhattan the high supply of cabs
enable a shorter waiting time, despite having a decentralized matching algorithm (street hailing) that is less efficient than an e-hail dispatcher.

The figure below shows the simulated histogram of waiting time distribution at each of the zones at 12 a.m. and an exponential probability density function with parameter empirical mean. As shown, the distributions are closely exponential. This justifies the imputation of taxi waiting time drawn from an exponential distribution in the demand estimation.

![Figure F.4: Distribution of waiting time at 12 a.m.](image)

F.2.2 Relation to the Cobb Douglas Matching Function

A parametric assumption made in the model is that the aggregate matching function at a location is follows a Cobb-Douglas functional form. To see if this functional form assumption is consistent with the microfoundation, I simulate a dataset for different number of cabs $C$, number of passengers $D$ and the resulting number of matches $m$ at each location in Manhattan and regressed $\log(m)$ on $\log(C)$, $\log(D)$ and a constant. Obtained a very high R-squared values for each of the regression, averaged at 0.9375, suggesting that the Cobb-Douglas matching function is a good approximation for the street hail taxi matching process.
F.3 Waiting Time Approximation Outside the Manhattan Grid World

Taxi zones outside uptown, midtown and Northern downtown Manhattan do not necessarily exhibit the uniform grid block structure, and therefore modeling cab movements and passenger arrivals on a matrix is feasible. The waiting times for these zones are imputed using the following approximation algorithm, make use of the New York City atomic map polygon project which allows us to identify each pickup location in terms of the street block it took place.

![New York city atomic map.](image)

**Figure F.5:** New York city atomic map.

As shown in the figure above, the partition is fine and I assume one passenger is waiting at each pickup. Suppose in a period of length $T$ passengers arrive uniformly and there were $n$ pickups where each of them occurred at time $t_1, t_2, ..., t_n$, with inter-arrival time $w_i$. By the assumption of uniform arrival rate the probability for a passenger to arrive between the $i$-th and $(i + 1)$-th pickup is $\frac{w_i}{T}$ and conditional on this, the mean waiting time is $\frac{w_i}{2}$ by the assumption of sole waiting passenger the expected waiting time is:
\[ \sum_{i=1}^{n} \frac{w_i}{T} \times \frac{w_i}{2} = \sum_{i=1}^{n} \frac{w_i^2}{2T} \quad (F.3) \]

I take \( T = 60 \) minutes. The figure below shows the estimated average waiting times other boroughs. Note that for waiting times in Bronx, Brooklyn and Queens include the passengers' expected waiting times for either a yellow or green cab. It is worth noticing that the evening peak effect are not prominent, as the shift change times for green cabs are less concentrated around evening, smoothing out the shift change supply shock.\(^1\) The dotted lines correspond to the average Uber waiting times. Outside Manhattan, the waiting time of Uber is significantly lower than taxicabs.

\[ \text{Figure F.6: Taxicab (yellow) and Uber (black) waiting times outside Manhattan.} \]

F.4 Waiting Time under a Centralized Dispatcher: Manhattan Grid World

I assume that under the centralized dispatcher, the passenger is matched with the closest available cab under the Manhattan distance. Under the map matrix the closest pair of waiting passenger and vacant cab is found using the \( k \)-dimensional (binary search) tree algorithm. I acknowledge that by minimizing the Manhattan distance between matched passengers and cabs the \( kd \)-tree algorithm does not necessarily also minimize the waiting time (because of road and traffic structures), but it is by far a relatively computation feasible search algorithm, with a time complexity of \( O(D \log_2 C) \).

\(^1\)This is also reflected in the number of active yellow and green cabs given in the 2016 TLC fact book.
Under the centralized dispatch platform, \( kd \)-tree search dispatch algorithm predicts the waiting time of a passenger would be the time required for the matched cab to travel the distance between the passenger and it. I acknowledge that under the Manhattan block matrix, the difference between displacement of passenger and cab and the distance travelled by cab to the passenger does not exceed 0.2 miles, so I add a random adjustment term which takes into account the additional time due to road feasibility. I also impose the restriction that the (Manhattan) distance between the matched passenger and the cab does not exceed 0.5 miles, as the TLC mandates e-hail dispatchers.

Assuming all cabs are dispatched under the same centralized planner, and the same number of cabs the figure below shows the updated waiting times.
Figure F.7: Simulated waiting times for taxicabs (solid) vs. Uber (dotted) in Manhattan under universal dispatcher.

As there are more yellow cabs than Uber cars in Manhattan at any given hour, under the same matching technology, cabs are able to offer a shorter waiting time for passengers almost anywhere and any time, except for around 5 p.m. at which time the number of active cabs decreases.
F.5 Waiting Time under a Centralized Dispatcher: Other Boroughs

The estimation of waiting time of cabs under a centralized matching regime poses a challenge: explicit simulations are not available because of map irregularities. Let the map space of a zone be given by $M \subset \mathbb{R}^2$ and define the Manhattan distance between two points $x$ and $a$ as $||x - a||_m$. Given the location of a passenger $a$ in $M$ and let there be $n$ random points drawn the set of cabs $C = \{x_i : i \in [1, N]\}$ from a distribution with bivariate CDF $F(x)$. Let the shortest distance between any cab to the passenger be given by $d(a; C) = \min_{x \in C} \{|x - a||m\}$. The cumulative distribution of $d(a; C)$ is:

$$H(\delta; a) = P(d(a; C) \leq \delta)$$

$$= 1 - P(d(a; C) > \delta) = 1 - P(||x_i - a||_m > \delta \ \forall i)$$

$$= 1 - \prod_i P(||x_i - a||_m > \delta)$$

$$= 1 - \left(1 - \int_{x:||x-a||_m<\delta} F(dx) \right)^n$$

Let the maximum distance between any 2 points on $M$ be $D$, the expected value of $d(a; C)$ is $\int_0^D \delta H(\delta; a)$. When the set of passengers are distributed according to CDF $G(a)$ the expected shortest distance between any cab and any passenger is $\int_M \int_0^D \delta H(\delta; a)G(da)$.\(^2\)

Intuitively, when $n$ approaches infinity, the expected shortest distance should diminish. For example, taking $M = [0, 1]$; when $F$ is uniform, for all values of $a$, $E[d(a; C)] = \frac{2}{n^2}$ or $\frac{2}{n(n+1)}$ depending on whether $n$ is odd or even, which converges

\(^2\)Technically, the sequence of the matching generates heterogeneity in $n$ for each passenger. I abstract away from the joint distribution of $n$ and $a$ by assuming that the entry rate of cabs cancels out the exit rate, so number of cabs is constant.
to zero at the limit. For a unit square map and uniform $F$, analytic solution of the integral yields the following the expected shortest distance for a passenger located at the center as a function as $n$:

![Figure F.8: Expected shortest distance.](image)

I acknowledge that it is not feasible to derive the analytic solution of the integral, due to irregularities of spatial distributions $F,G$, map domain $M$ and subdomain $\{x : x \in M & ||x - a||_m < \delta\}$. Given the complexity of the problem, Monte Carlo simulated evaluation of the expected shortest path is infeasible, as well. Discretizing the integral $\int_{x : ||x - a||_m < \delta} F(dx)$ as a weighted sum of product of density $f$ and the area in the subdomain $\{x : ||x - a||_m < \delta\}$, I transform the map of the borough into a polygon under pixel map and a sparser grid point, for each grid point $x$ is assigned the heat level $\tilde{F}(x)$, which is the empirical street hail density over the map. The approximated cumulative probability for $d$ is: $\tilde{H}(\delta) = 1 - (1 - \sum_{x : ||x - a||_m < \delta} \tilde{F}(x))^n$. 

154
I assume that the spatial distribution of passengers $G$ is identical to that of cabs, $F$: either through learning by doing or dispatcher app info\textsuperscript{3}, drivers eventually learn where passengers are typically located and the probability measure in these zones become denser. As expected, supply of cabs are higher than Uber and less elastic to time, under the regime that all cabs are operated through a centralized dispatcher, the waiting times for cabs manage to become shorter than the average Uber waiting time in the same borough.

\textsuperscript{3}For example, Uber's driver app provides a heatmap to drivers to notify them about demand surges.
Table F.1: Waiting times outside Manhattan.

<table>
<thead>
<tr>
<th></th>
<th>full cab dispatch</th>
<th>Uber average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Manhattan</td>
<td>2.78</td>
<td>3.88</td>
</tr>
<tr>
<td>Bronx</td>
<td>4.51</td>
<td>6.53</td>
</tr>
<tr>
<td>Brooklyn</td>
<td>3.94</td>
<td>4.53</td>
</tr>
<tr>
<td>Queens</td>
<td>3.83</td>
<td>5.38</td>
</tr>
</tbody>
</table>
E-hail and other digitization devices are also seen as a channel to notify drivers about real time changes in demand patterns, in particular under unanticipated shocks. In this section of the appendix, I study the allocation effects of surge pricing under the presence of an unanticipated shock from forward looking drivers’ perspective. Realistically, this requires drivers to make decisions in a very short time frame and the actual decision model for drivers in short run possibly deviates from the rational expectation constraint of competitive equilibria, as rational expectation of other drivers’ and riders’ behavior is typically a product of learning in the long run (Haggag, McManu & Paci, 2017). Nonetheless, I present the simulation result in which when the shock is applied to the system, drivers take the status quo of cab distribution and history of matching as given and update their beliefs as if the rest of the game is played from this point until the end of time horizon.

I apply a demand shock of 1,000 riders in Midtown West that sustains for two periods. Fig. G.1 compares the supply and demand under surge pricing and belief updating with (a). no surge pricing and drivers are not aware of the shock, (b). no surge pricing, but drivers have anticipated the shock and are able to reoptimize.

Fig. G.1 capture the effect of surge pricing on filtering excess demand (a) to ensure more matches made under sufficiently high request fulfillment rates, plus the effect on alerting drivers about the demand shock, a feature that comes with digitized dispatch platforms (b).
**Table G.1:** Equilibrium outcome.

<table>
<thead>
<tr>
<th></th>
<th>Surge pricing</th>
<th>No surge, unanticipated</th>
<th>No surge, anticipated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand</td>
<td>6,850</td>
<td>7,039</td>
<td>7,044</td>
</tr>
<tr>
<td>Matches</td>
<td>6,672</td>
<td>6,658</td>
<td>6,663</td>
</tr>
<tr>
<td>Pickup prob</td>
<td>0.9587</td>
<td>0.9523</td>
<td>0.9548</td>
</tr>
<tr>
<td>Fulfillment rate</td>
<td>0.974</td>
<td>0.9459</td>
<td>0.9459</td>
</tr>
<tr>
<td>Waiting time</td>
<td>4.6269</td>
<td>5.108</td>
<td>5.1576</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$39,183</td>
<td>$40,699</td>
<td>$39,292</td>
</tr>
<tr>
<td>Platform earnings</td>
<td>$84,121</td>
<td>$80,224</td>
<td>$80,464</td>
</tr>
<tr>
<td>(\Psi)</td>
<td>6.1954</td>
<td>44.22</td>
<td>45.26</td>
</tr>
</tbody>
</table>

**Figure G.1:** Supply and demand under a shock.
Bibliography


159


