MODELING EMERGING APP-BASED TAXI SERVICES: INTERACTIONS OF DEMAND AND SUPPLY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Wenbo Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2018

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Satish V. Ukkusuri, Chair

Lyles School of Civil Engineering

Dr. Hua Cai

School of Industrial Engineering

Dr. Konstantina Gkritza

Lyles School of Civil Engineering

Dr. Harsha Honnappa

School of Industrial Engineering

Dr. Shreyas Sundaram

School of Electrical and Computer Engineering

Approved by:

Dr. Dulcy M. Abraham

Head of the School Graduate Program

For their endless support and love, dedicated to my respected parents.

ACKNOWLEDGMENTS

First, I would like to express sincere gratitude to my dissertation advisor Dr. Satish V. Ukkusuri for all guidance, advices, supports, and encourages in both personal and professional activities. The flexibility in conducting research, as well as open mind, helped me to investigate research problems beyond transportation, broad my knowledge, and raise my creative and independent thinking.

Second, I would like to thank all committee members: Dr. Hua Cai, Dr. Konstantina Gkritza, Dr. Harsha Honnappa, and Dr. Shreyas Sundaram. I appreciate their time for reviewing my work and value their ideas and insightful comments at different stages. In addition, special thanks to Dr. Harsha Honnappa for his important guidance in Chapter 7.

Third, I would like to thank for the financial supports from China Scholarship Council during my first four-year study, including tuition, stipend, and round-trip flight tickets to US.

Last, I would like to thank all lab members for wonderful experiences and meaningful ideas. It's so lucky for me to have colleagues and friends: Dheeraj, Xinwu, Tho, Hemant, Fasil, Nishtha, Taka, Maria, Zengxiang, Lu, and Jiawei. Working as a research assistant in Urban Mobility and Network Intelligence (UMNI) Lab, as well as with nice lab members, are really helpful and has been one of rewarding experience in my career.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Zhang, Wenbo Ph.D., Purdue University, December 2018. Modeling Emerging App-Based Taxi Services: Interactions of Demand and Supply. Major Professor: Prof. Satish V. Ukkusuri.

The app-based taxi services (ATS) has disrupted the traditional (street-hailing) taxi services (TTS) leading to transformative changes in the urban taxi markets and its impacts on mobility, design and environment. However, the current modeling of these new mobility markets is limited in its understanding of: (1) the underlying factors that influence the growth of the ATS market; (2) the competition of ATS and TTS markets; (3) pricing in the ATS market; (4) system wide tools to understand the impacts of the market. The overarching goal of this dissertation is to address four fundamental processes of taxi system, ranging from demand generation, supply generation and exiting, dynamic pricing generation, and vehicle-passenger matching over road network. This dissertation achieves these goals by using original large scale datasets to characterize disruptive changes in mobility, understand strategic behaviors of stakeholders, and formulate system dynamics.

This dissertation develops various modeling structures and estimation methods, motivated from statistical, econometric, machine learning, and stochastic approaches. First, we adapt multiple econometric models for demand, supply, and platform-exiting (offline) behaviors, including mixture model of spatial lag and Poisson regression and mixture model of spatial lag and panel regression. It is apparent that all proposed econometric models should be corrected with spatial lag due to significant spatial autocorrelations. The results indicate effectiveness of dynamic pricing in controlling demand, however, it also shows no impacts on driver's online and offline behaviors. Then a dynamic pricing generation problem is formulated with multi-class classifica-

tion. This model is empirically validated for the impacts of demand and supply in dynamic price generation and the significant spatial and temporal heterogeneity. Last, we propose a queueing network consisting of taxi service queues for vehicle-passenger matching and road service queue for vehicle movements at homogeneous spatial units. The method captures stochasticity in vehicle-passenger matching process, and more importantly, formulates the interactions with urban road traffic.

In summary, this dissertation provides a holistic understanding of fundamental processes that govern the rapid rise in ATS markets and in developing quantitative tools for the system wide impacts of this evolving taxi markets. Taken together, these tools are transformative and useful for city agencies to make various decisions in the smart mobility landscape.

1. INTRODUCTION

1.1 Background

1.1.1 The focus of the dissertation

Taxis are necessary and important component of urban mobility. Figure 1.1 summarizes the share of multiple urban transportation modes in New York City (NYC), one of the largest metropolitan areas in the world. It is straightforward that around 10% of urban trips are undertaken by taxis or similar services. Although taxi services have much fewer rides than bus and metro, it provides 24/7 door-to-door mobility service with superior conveniences that the mass transit system cannot match and substitute. As the rise of information and communication techniques, the smart urban mobility is also reshaping the traditional taxi services (TTS) through taking advantages of smart-phones and location aware techniques, leading to a new service type of the app-based taxi services (ATS). The ATS (also called mobility-on-demand, ride-sourcing, e-hailing, etc.) introduces smart-phone based applications for both customers and drivers, who can log on to their client applications for real time updates on surrounding system dynamics (e.g. available drivers/customers and fare rates). It has been succeeding in attracting users and been getting explosive growth after the launch year of 2010, in particular in recent years. The two most representative transportation network companies (TNCs), Uber and Lyft, are operating in more than 360 and 190 cities, respectively. Taking the example of NYC, the five largest TNCs, including Uber, Lyft, Via, Juno, and Gett, are dispatching more rides than the TTS (i.e. NYC yellow and Boro taxicabs), and since early 2015 have been observing more active ATS driver partners, as shown in Figure 1.2(a) and 1.2(b). Compared with the traditional NYC yellow taxis that have operated for decades, only the largest TNC, Uber, has almost same ridership and around four times as many registered vehicles as yellow taxicabs, just after five years' expansion. In addition, the ATS shows significant improvements in efficiency, as shown by the much lower service hours, larger number of dispatched vehicles, and almost the same ridership in Figure 1.2(c). Beyond the taxi market, the ATS are also verified to be competitive with transit and other private transportation modes. In San Francisco, around 75% of casual carpool users were previously public transit riders and over 10% formerly drove alone [4, 5]. It is anticipated that the new business model together with other similar shared mobility services, will continue to grow and further disrupt (or even reshape) urban mobility. Thus, **the focus of this dissertation is on the emerging taxi market comprising of both ATS and TTS, as well as interactions**.

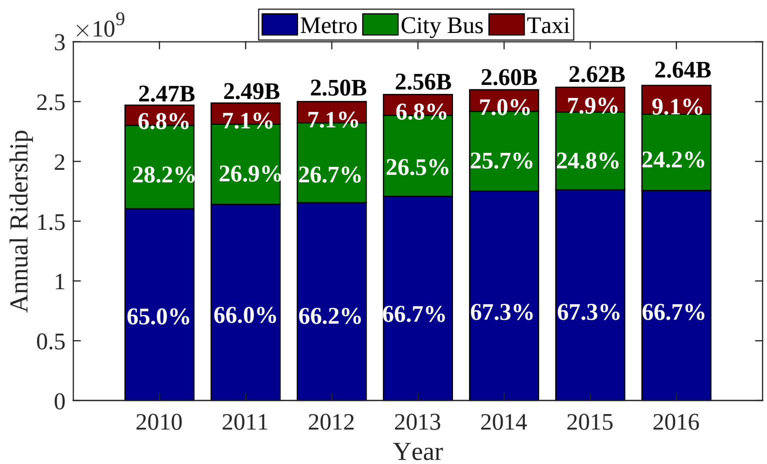


Fig. 1.1.: Annual ridership distribution by transportation modes

B-Billions; Data Sources: NYC Taxi and Limousine Commission & NYC Metropolitan Transportation Authority

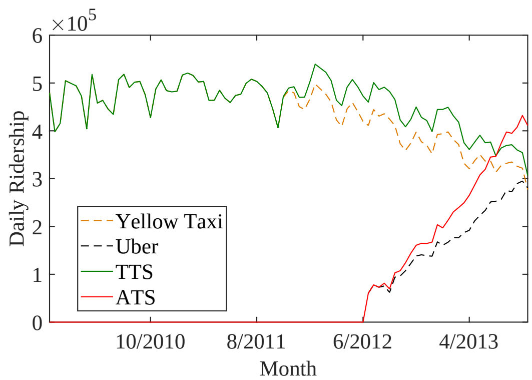
1.1.2 Overarching goals

Underneath the smart-phone based applications that connect customers, drivers, and service providers, ATS are also transforming customer expectations and disrupt-ing the way we move around the city. The disruptive changes by ATS are summarized as follows:

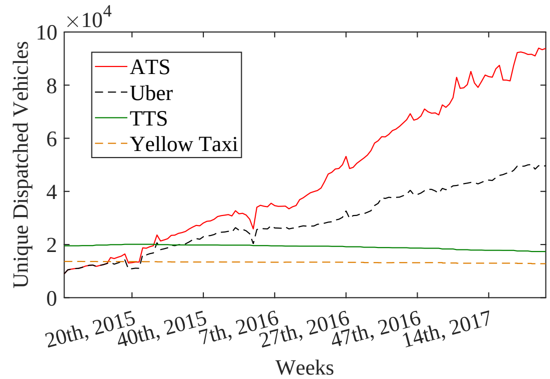
- *Supply (Drivers):* Without regulations or permits (e.g. medallions), any drivers who own a private car can apply for a license with much lower costs to serve as ATS drivers. Moreover, in contrast to the minimum eight-hour shift required by TTS, ATS drivers can go online and offline anytime and anywhere. Thus, they are able to work part-time for just few hours per week, such as early morning, late night, and commuting to home/work. According to one study on Uber's driver partners [6], more than 50% of their drivers work less than 15 hours per week, while no more than 10% work more than 50 hours per week.

- *Demand (Customers):* Obviously, the ATS is designed for smart-phone users. You can request and pay for rides with smart-phone based applications. Although ATS platforms provide website versions, they are much less popular than the smart-phone usage. On the other hand, the smart-phone based services are limited to the potential customers who must own a smart-phone. However, free entry regulations enable ATS providers to attract various types of vehicles, design differentiated products, and thus meet diverse travel demand than TTS. For example, UberPool is the economic ride sharing service; UberX and UberXL are economic services; UberFamily and UberWAV are designed for families with children and the disabled, respectively; and UberBlack and UberSUV are premium services.

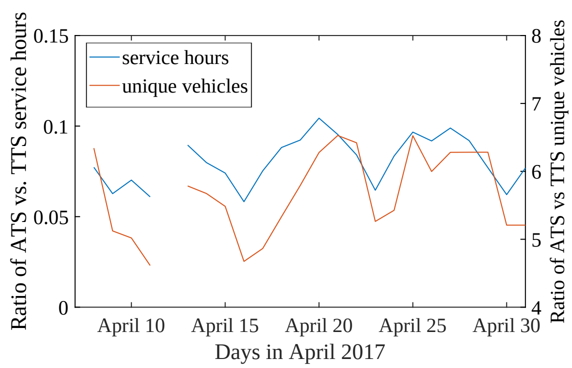
- *Smart (Dynamic) Management:* The smart-phone usage also enables TNCs to collect demand and supply dynamics in a timely manner, measure system demand and supply mismatch, and develop instant responses to any demand-supply imbalance. One strategy that is representative of smart management is dynamic pricing, which is also called surge pricing. The idea behind surge pricing is to incentivize additional drivers to serve an area experiencing high



(a) Daily ridership by different taxi services



(b) Weekly dispatched vehicles



(c) Service hours and daily unique vehicles

Fig. 1.2.: TTS vs. ATS operations in NYC

Data Sources: NYC Taxi and Limousine Commission and collected dataset in April 2017 demand, and to reduce demand at least temporarily, with higher fare rates. Such spatially and temporally differentiated policy is thought to be effective in maintaining demand-supply equilibrium [7]. The real-time control strategy is very different to having fixed rate for all times across the whole city by TTS. Another disruptive change in management is user centered. With access to a huge number of observations on riders' requests and available drivers, the TNCs have a deep understanding of users (both riders and drivers), including what they need, what they value, their availabilities, and also their limitations. This understanding can assist in the development of new strategies and improve user experiences, such as the pilot program of Uber's route-based pricing that charges customers based on what it predicts they are willing to pay. This is different from a decentralized system depending on individual's experiences and knowledge.

- *Monopolistic Competition Taxi Market:* The rise of ATS has reshaped the monopolistic taxi market dominated by TTS, and is transforming the urban taxi market into a new market structure of monopolistic competition. In the past few decades, most taxi systems have been strictly regulated by city agencies in terms of both fare rates and fleet size. There are no close substitutes for door-to-door urban mobility services, and it was impossible to join the market without permits from city agencies. However, the TNCs are breaking down these barriers and transforming the market into monopolistic competition by providing the similar door-to-door urban mobility services but by way of e-hailing. The significant market share of a few giant TNCs also prevents the development or even entry of other small or local TNCs. Within the market transformation, the complexities and difficulties are challenging to understand the emerging taxi market, since there are many more interactions among different services, as well as stochasticity within each service.

The urban transportation system should offer a reliable, safe, accessible, and comprehensive mobility service that promotes the public good and meets the needs of residents and visitors. Face with the rapid growth of ATS and the corresponding significant disruptive changes to the urban transportation system, city management agencies have sought to answer fundamental questions on the potential impacts of these changes: measuring urban mobility option transformations, managing efficient movement, supporting a comprehensive and sustainable transportation system, promoting equitable growth, and ensuring safety [7]. Without a better understanding of the urban mobility transformations and actions that contribute to these challenges, cities and communities are subject to potentially catastrophic outcomes in health and welfare.

Unfortunately, we still do not have clear, quantitative, and reliable methods to determine these impacts, thus there is a lack of transparency of complex taxi systems with multiple service providers, as well as sustainable control policies for the emerging taxi market. There are several barriers preventing in-depth discussions. **The first barrier** is the availability of high resolution mobility data that supports a clear understanding of how the taxi system is transforming. The recent introduction of mobility data such as a time-stamped trajectory has enabled the rules and behaviors of traditional taxicabs to be uncovered, using taxi ride distribution and pattern recognition, demand generation and mode choice, empty taxicab searching, and driver-passenger matching. However, it is difficult to acquire such dataset for ATS due to privacy and commercial concerns. The only officially released data are the administrative dataset for the public and the aggregated performance dataset for the city agencies. Detailed representation of urban dynamics and individual level patterns are difficult to discern with this type of data and these methods. **The second barrier** lies in the transformations and flexibility of user behaviors. There are currently limited methodologies that allow us to analyze ATS ride generation and mode choice among different services, as well as ATS drivers' responses (online/offline behaviors) to market dynamics, which are completely new topics in the emerging taxi market, compared to previously fixed supply shift and monopolistic service of the TTS. More importantly, the underlying mechanism of dynamic pricing generation and its impacts on demand and supply are also understudied. The opacity of interactions among pricing, demand, and supply prevents the accurate simulation and modeling of the emerging taxi market. **The last barrier** is the complex interactions in the emerging taxi market. The monopolistic competition taxi market results in a high degree of competition among similar service providers. Only very few are focusing on the competition and their modeling from the past few years, but these findings are not well validated with real examples. Moreover, most of the past and ongoing studies ignore the congestion externalities. Hence, this dissertation proposes a comprehensive modeling structure for the emerging taxi market which takes into account both service competition and congestion externalities.

Focusing on the ATS and TTS, this dissertation presents a comprehensive discussion on stakeholders' behaviors and their interactions. The goal of this dissertation is to *develop the data analytics motivated from statistical, econometric, machine learning, and stochastic approaches to understand disruptive changes in urban mobility induced by the rise of ATS. For example, we seek to explore the interactions among demand, supply, dynamic pricing, and the urban road network.* In particular, we combine the novel mobility dataset for ATS with the current available TTS trip record data to recognize disruptive changes in urban mobility at the aggregate level. Using the novel mobility dataset, we develop statistical estimation techniques for understanding stakeholders' strategic behaviors, including passenger generation, ATS drivers' offline, and dynamic pricing generation. In addition, we model the monopolistic competition taxi system with both ATS and TTS, and quantify passenger-vehicle bilateral searching, as well as interactions between the taxi system and urban road network. The realization of this goal will provide urban transportation agencies with reliable and accurate tools for characterizing disruptive changes induced by ATS, yielding better policy making, and creating a sustainable and smart urban mobility system.

1.2 Motivations

1.2.1 The rise of smart-phone and app-based taxi services

Of the various technologies impacting our societies, it appears the smart-phone is more popular and having a more profound effect in many contexts. It is reported that there will be 2.1 billion smart-phone users (28% of the total population) in 2016, which will be tripled to more than 5 billion (65% of the total population) by 2019 [8]. A survey from Pew Research Center in 2015 found that advanced economies have considerably higher rates of smart-phone adoption, such as 77% of adults in U.S., 72% of adults in U.K., 88% of adults in South Korea, and 68% of adults in China [9]. These smart-phones are typically equipped with perfect hardwares, for instance fast wireless communication and location awareness, as well as applications that can enable users to access the Internet and a variety of nontraditional phone activities such as, travel, reading, and job searching. They are certainly not limited to just calling or texting.

The ubiquitous use of smart-phone, as well as the changing ways of using smart-phones, promoted the rise of technology-enabled mobility services and travel apps that allow users to hail a ride by utilizing smart-phone based applications, or to receive real-time operation information. It is reported that there are presently more than 300,000 travel apps to address almost any situation and desire and around 75% of adults access travel apps during travel, just less than the 89% taking pictures and 85% access social media [10]. The survey conducted in 2015 by Pew Research Center also found that about 15% of American adults have used ride-hailing apps and half of all Americans are familiar with these services but have not actually used them. In addition, there are also higher rates of usage in some special groups; for instance, 28% of 18- to 29-year olds, 19% of 30-to 49-year-olds, 29% of college graduates, and 26% of high income households (over $75,000) have used these services [11]. On the other hand, the TNC experiences an amazing increase in market share in just few years after launching around 2010. Findings by the Benenson Strategy Group show that Uber had over 160,000 active drivers by the end of 2014 in the US, receiving total payments of $656.8M during Q4 2014 from the company and completing 172 million rides [6]. By 2017, the total number of Uber drivers and ridership worldwide are expected to be more than 1.5 million and 2 billion, respectively. In addition, the market share of the TNC has slightly passed TTS, even with fewer service hours, as shown in Figure 1.2.

1.2.2 Taxi system modeling with differentiated service providers

In the monopolistic taxi market only dominated by TTS, the system dynamics are primarily described by bilateral passenger-taxicab searching. The taxicabs are immediately available for other passengers after one drop-off, and then decide searching routes over the road network. The passengers wait at the curbside or move around to catch available taxicabs. The bilateral searching behaviors play important roles in the system performance; for instance, waiting/searching time and utilization. However, the rise of the ATS disrupted the system and introduce more complexity and stochasticity, as shown in Figure 1.3. The first change was the competition for passengers between the ATS and TTS. The second change was the ATS drivers' decisions on online/offline due to flexible working hours. Moreover, the ATS platform examines demand and supply frequently, and utilizes dynamic pricing to seek a balance in demand and supply. Thus, the emerging monopolistic competition taxi system involves not only bilateral passenger-vehicle searching but also ATS drivers' and passengers' strategic behaviors.

The academia has a long standing interest in monopolistic taxi system modeling and has derived various methods at both the aggregated and disaggregated levels.

- The first line of research is the aggregated models that formulate the relationships among system performance metrics; for instance, nonlinear simultaneous equations of system performance [12], queue theory [13, 14], and neural network [15]. The major limitation of aggregated models is addressing the stochastic nature of components and spatial effects.

- The second line of research is the equilibrium models that investigate dynamics of drivers and passengers, and take into account internal and external factors. Based on different definitions of equilibrium, there are three modeling structures: demand-supply equilibrium [16,17], competitive equilibrium [18,19], and other defined stationary distribution states [20,21]. However, almost all equilibriums are built on weak assumptions of passenger-driver matching, thereby leading to reliable estimations of utilities and state distributions. Moreover, the supply-demand equilibrium is mainly designed for a perfect market with perfect information, homogeneous products, no barriers to entry, and profit maximization of service providers. Obviously, we cannot model the TTS or the emerging monopolistic competition market as perfect one.

- The last line of research generally emulates individual behaviors and interactions with others by representing system participants as agents or nodes in a graph; for instance, using graph theory [22] and agent-based simulation [23]. Obviously, the large-scale system will include a considerable number of agents or nodes, as well as their interactions. Additional methods should be introduced to address heavy computational burdens.

Besides the aforementioned limitations, we have not seen an elaborate system modeling of emerging monopolistic competition taxi market with both ATS and TTS, except for limited extensions from equilibrium models. The studies [24,25] have thus far applied the demand-supply equilibrium model to the scenario of such emerging taxi market, since it still assumed a static pricing scheme rather than the dynamic one utilized by the ATS. Although other studies [26] considered dynamic pricing for the ATS under a framework of competitive equilibrium between the ATS and TTS, they have failed to precisely formulate vehicle-passenger matching, as well as passengers' and drivers' strategic behaviors.

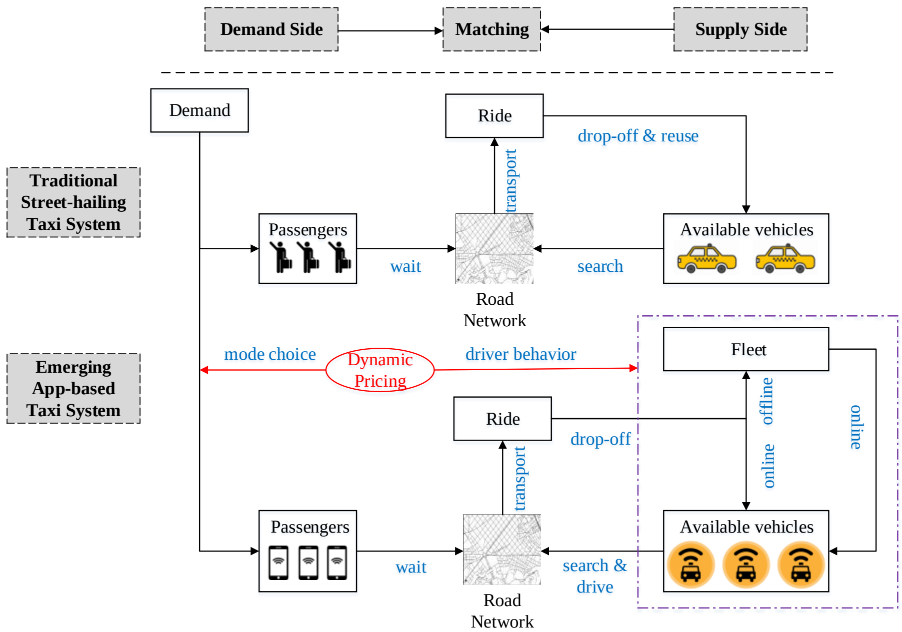


Fig. 1.3.: Illustrations of urban taxi system with multiple services

To sum up, there remains a lack of elegant system modeling structures for such emerging taxi market, in particular for individual behaviors under dynamic pricing. Here, an elegant system modeling should not only captures the dynamics and competition in both taxi services, but also investigate the passengers' and drivers' movements across heterogeneous spatial units. This also leads to another key concern in congestion externalities, indicating the interactions between taxis and regular flows of the urban road network. At this point, the first gap in the literature is identified as follows.

- **Gap 1**: *There is no one elegant system modeling in the literature, accurately describing dynamics of taxi stakeholders and external interactions with the urban road network.*

Besides an overall modeling structure, the system modeling is developed to describe one key process that is matching between passengers and empty vehicles. The modeling outputs are passenger waiting time, driver searching time, and vehicle utilization, which are critical for efficiency measurement and system optimization. One common solution approach in the literature is Cobb-Douglas production function, built upon the meeting rate derived from number of passengers and number of available drivers [14, 17]. However, it is hard to accurately calibrate related parameters and explain the nature of those parameters. Another solution is based on the heuristic assumption of an even distribution of drivers and passengers within one specific area, and estimates waiting and searching time considering area characteristics [26]. Similar to the Cobb-Douglas production function, this assumption has been questioned recently and is without empirical evidence. Additional studies have also implemented econometric analyses on drivers' searching times with duration models and exogenous variables, but without any representations of passenger-vehicle matching [27,28]. Since these solutions are inefficient in describing the demand-supply matches, the next gap can be stated as follows.

- **Gap 2**: *There is no an aggregate system modeling, describing the stochastic nature of demand-supply bilateral search and leading to reliable system performance measurements.*

1.2.3 Interactions among demand, supply, and dynamic pricing

Another question of interest within the emerging taxi market should be the strategic behaviors of passengers, drivers, and centralized platforms. In specifics, these strategic behaviors include the demand generation of both ATS and TTS, ATS supply under flexible working hours, and dynamic pricing generation controlled by ATS platforms. More importantly, the passengers' and drivers' strategic behaviors are more special (different from behaviors under fixed fare rate) and interesting under dynamic pricing, which have not been extensively discussed in the literature.

Demand generation is a long standing topic in taxi studies. However, limited by the data availability, demand refers to the number of rides from one location over a defined period in most studies. The latest progresses are multivariate and generalized regression, while understanding demand generation with exogenous variables. The methods generally take the ridership as the dependent variable and multiple other external factors as the independent variables, such as socioeconomics, land use, and built environment [29, 30]. Cohen et al. also considered the modeling structure as a tool for the UberX purchase rate under surge pricing. Given the facts of excessive zero ridership in many zones and the count nature of ridership, Zhang et al. introduced the zero-inflated negative binomial model [31]. However, another key feature of taxi system is spatial autocorrelation, which can not be well addressed by traditional econometric approaches. Since most econometric approaches in the literature are built upon the assumption of normal distribution for demand. Alternatively, spatial econometrics are utilized to capture spatial heterogeneity and autocorrelation; for instance, spatial lag, spatial error, and geographically weighted regression [32,33]. In final, the generalized forms are also utilized to relate taxi demand to endogenous factors; for instance, waiting time and fares. To sum up, although new solutions to taxi ridership specifications have emerged, the impacts of emerging ATS and surge pricing on both ATS and TTS ridership remain unclear. Moreover, the current modeling structure cannot address the modal correlations among multiple taxi service types very well, although the spatial autocorrelation is discussed in recent years.

On the other hand, several studies have also explored the fleet size and its influencing factors. Similar to the ridership analyses, the number of vehicles in the system is modeled under a regression structure. Schaller and Kamga et al. identified the influencing factors for the TTS fleet [34]. Yang et al. also proposed a neural network theory to address the nonlinear relationship between the taxi fleet and external factors. Moreover, the ATS fleet is also examined for potential influencing factors with econometric models. For example, Qian and Ukkusuri relate new ATS drivers to surge pricing and current demand in the system [26]. Obviously, there are much fewer studies on supply and influencing factors, indicating there is a lack of in-depth discussions compared to the demand generation problem. Again, the existing simple models cannot address those problems identified in ridership specifications, such as spatial-temporal effects and excessive zeros.

Considering the limitations discussed in the previous paragraphs, the next gap can be stated as follows

- **Gap 3**: *There is no well-developed models for demand and supply generation considering the existence of ATS, as well as the emerging complex relationship at both spatial, temporal, and modal scales.*

The second interesting topic in the monopolistic competition taxi market is the behavioral decisions by drivers. One vehicle becomes empty after drop-off and searches for next potential passengers if it remains online. The process contains a set of driver decisions regarding searching routes or destination choices. Moreover, the flexible working hours enable ATS drivers to go offline, leading to stop-working decisions.

Such discrete choice problems are generally formulated with expected utility theory and the logit modeling structure, under the assumption of rational behaviors for maximum benefits. In terms of the drivers' customer searching decisions, almost all existing studies refer to the logit based modeling structure, although a few extend to a multi-stage approach that can capture dynamics along the route instead of including variables only at origins and destinations [35–37]. On the other hand, very few studies have investigated ATS offline behaviors with logit-based modeling [38]. However, the expected utility theory, together with logit-based modeling, are not well validated with empirical cases. More importantly, the current progress in behavioral economics is challenging the expected utility theory, because of rational behavior assumptions and the lack of loss aversion.

Hence, we justify a new gap in the literature for modeling behavioral decisions by drivers and passengers.

- **Gap 4**: *There are no detail and accurate descriptions of ATS driver decisions with appropriate behavioral modeling and empirical validations.*

Last but not the least, we shed light on the dynamic pricing that is one representative feature of ATS. However, the existing literature generally assumes the fixed price rate to simplify problems. This contradicts reality in the ATS. With limited empirical datasets, very few studies are exploring this interesting characteristic and how it is generated based on demand and supply. Guo et al. utilized a curve fitting technique to identify surge pricing generation from ratio of successful orders [39]. Chen et al. proposed a multivariate regression for surge pricing with the independent variables of supply-demand difference, waiting time, and current level of surge pricing [40]. However, these models did not effectively figure out how ATS platforms generate surge pricing. Moreover, they did not consider any potential spatial and temporal effects. The above concerns led to another gap as follows.

- **Gap 5**: *There are no formulations describing dynamic pricing generation with demand and supply dynamics.*

1.2.4 Challenges of large-scale observations on app-based taxi services

To empirically validate the proposed modeling, we should utilize a big mobility dataset with large-scale observations on occupied trips and empty vehicle movements in one city or system. There are three typical types of datasets for analyses.

The first is survey or interview data, including stated preferences in behaviors and individual socioeconomic background. The stated preference surveys can be made for passengers and explore their preferences in using ATS. Shaheen et al. and Rayle et al. collected 503 and 380 questionnaires in San Francisco [4,5], respectively. Another similar survey was made for 230 residents in Beijing [41]. Such surveys can be also designed for drivers and investigate their preferences on serving the ATS. A web survey of Uber's driver partners was completed in December 2014 and November 2015, and obtained a dataset featured 601 and 833 drivers in 2014 and 2015, respectively [6]. In addition, Schwieterman and Michel took 50 real trips by UberPool and Chicago metro respectively, and recorded the travel performance [42]. The type of datasets can uncover individual's preferences on choices relating to socioeconomic background. However, it is difficult to extend the dataset to a larger group of passengers, drivers, and rides.

An efficient alternative to manual surveys are the datasets directly requested from TNCs or other agencies. Dias et al. requested data derived from the 2015 household travel survey of the Puget Sound Regional Travel Study, including attitudes and values, technology ownership and usage, membership in and use of car-sharing and ride-sourcing services, and future intended adoption and use of autonomous vehicle technologies [43]. FiveThirtyEight obtained 4.5 million Uber pickups in NYC from April to September 2014 and 14.3 million additional Uber pickups from January to June 2015. All these Uber pickups are shared by the NYC Taxi & Limousine Commission (TLC) on July 20, 2015, under Freedom of Information Law [44, 45]. Cramer and Krueger collaborated with the Uber research staff, who can provide statistics on the fraction of occupied time and capacity utilization rate based on Uber's administrative database, for 2000 Uber drivers in five different cities [46]. Cohen et al. also worked with Uber to collect approximately 54 million UberX sessions in Uber's four largest U.S. markets (San Francisco, New York City, Chicago, and Los Angeles) over the period of January 1, 2015 to June 17, 2015. For each session, they observed the surge price, underlying surge generator (which the consumer does not see), Uber-defined geographic region, time and date, anonymized rider ID, Uber's prediction of expected wait time (consumers see wait time in minutes), product, and the ultimate decision of the rider whether to request a car [47]. Chen et al. extracted 251,344 and 1,678,289 (about 20% of all trips) completed trips from DiDi between September 7 and 13, 2015, and between November 1 and 30, 2015 in Hangzhou, China. An individual trip record includes the pickup and drop-off location and time, order ID, passenger ID, passenger order time, driver response time, trip distance, type of service, car level, trip cost, and whether to share rides [48]. Although the collaborations with agencies can allow much more observations, the current available datasets shed light on the passenger or demand side. There are no any information on partner drivers, such as when and where they go online/offline and where they move for next passengers.

Last, the emerging datasets relating the ATS leverage current progress in information techniques, such as developer application programming interfaces (APIs) and scraping, to capture information as needed. Salnikov et al. developed a price comparison application between the ATS and TTS in NYC based on Uber's API and collected a total of 25,804 queries [49]. Hughes and MacKenzie collected estimated wait times for an UberX vehicle through Uber's API over two months in 2015 in Seattle [50]. Guo et al. included a script in the platform of Shenzhou Ucar and acquired a dataset containing the event code (i.e., EstimateFee or Create Order), the IMEI of the passenger's device, location information, the estimated trip fare from EstimateFee, the order's boarding/arriving locations and time, the unique ID for the user (and the driver, the car, etc.), the type of order, the amount of money, the corresponding dynamic pricing multiplier, and so on, from November 2015 to March 2016 in Beijing [39, 51]. Chen et al. collected data from 43 emulated clients placed in midtown Manhattan and downtown San Francisco between April 3rd to 17th (391 GB of data and 9.3 million samples) and April 18th to May 2nd (605 GB of data and 9.4 million samples), respectively. Each client captured nearby vehicle trajectory, waiting time, and surge pricing [40]. These datasets can include more detail of vehicle movements and system performance that are not released to collaborators by agencies. However, the current datasets are still limited to a small scale, other than the whole system.

After checking all available datasets, we identified a new gap in data availability that is likely ignored in most studies. A high-resolution dataset could present solid empirical evidences of our proposed methods.

- **Gap 7**: *There are almost no large-scale mobility datasets on representative ATS products, including driver, passenger, and platform dynamics over perfect spatial and temporal resolution.*

1.3 Thesis Objectives

The main objective of this dissertation is to narrow the gaps in the literature by providing innovative research ideas related to fundamental understandings of the monopolistic competition taxi market with both the ATS and TTS. The specific objectives are presented as follows:

- Collect a large-scale mobility dataset for one representative ATS platform that contains high-resolution information on vehicle and passenger movements, as well as real-time fare rate.

- Characterize the spatiotemporal variations in passenger and empty vehicle flows of the ATS versus TTS at the aggregate level. The specific questions relating to the objective are listed as follows:

  Q1: *Are the released datasets sufficient for solving our problem? If not, how can we obtain TTS and ATS operation data?*

  Q2: *What are the appropriate spatial and temporal scales for our problem?*

- Develop a demand and supply generation model for ATS and TTS taking into account surge pricing. The specific questions relating to the objective are listed as follows

  Q3: *What are the impacts of dynamic pricing on the ATS, TTS, or both ridership?*

  Q4: *What exogenous variables are influential for demand generation?*

  Q5: *What are the impacts of dynamic pricing, as well as exogenous variables, on ATS supply generation?*

  Q6: *What modeling structures are appropriate for such spatiotemporal datasets?*

  Q7: *Are there correlations occurring while the ATS and TTS rides generate? If so, how should we adjust the modeling structure?*

- Develop a decision model for ATS drivers that can describe their aggregated platform-exiting behaviors.

  Q8: *What internal factors contribute to drivers' choices in exiting ATS platform?*

  Q9: *Are there differences in exit probabilities across aggregated zones? If yes, what variables are relevant?*

  Q10: *What modeling structures are appropriate for modeling such aggregated decisions?*

- Develop a model to investigate dynamic pricing generation, considering demand, supply, and potential spatio-temporal effects.

  Q11: *What internal factors contribute to dynamic pricing generation?*

  Q12: *Does spatio-temporal heterogeneity exist while generating dynamic pricing?*

  Q13: *Which modeling structures perform better?*

- Devise a unified modeling framework that not only describes the stochastic nature of passenger-vehicle matching, but also captures the traffic state on the urban road network.

  Q14: *How do drivers meet passengers within an aggregated region?*

  Q15: *How do taxi vehicle movements interact with regular traffic flow over the urban road network?*

1.4 Contributions

This dissertation investigates the potentials of leveraging high-resolution mobility datasets to explore strategic behaviors of passengers, drivers, and centralized platforms, and proposes a unified modeling framework for taxi stochastic processes with significant implications for impact analysis and regulation development. It contributes to the field of ATS and TTS studies in the following ways:

- **Use of a novel mobility dataset for the ATS**, as shown in Figure 1.4. This dissertation utilizes a new data source that contains high-resolution mobility information from ATS drivers and passengers, as well as platform's dynamics, obtained through scraping the ATS platforms. Moreover, this dataset has a good spatial and temporal coverage in NYC. It would be impossible to collect a dataset with such coverage (high frequency trajectory from hundreds of thousands of vehicles within NYC for one month) if using traditional collection methods or directly requesting the data from TNCs. This large-scale mobility dataset, together with the publicly available mobility dataset for TTS, presents promising opportunities for taxi system modeling and enhancement which can not be adequately explored with limited data availability.

- **Understanding strategic behaviors of the stakeholders**, as shown in Figure 1.5. It provides detailed mathematical forms of strategic behaviors of passengers, drivers, and platforms; for instance, demand generation and competition, ATS supply and drivers' behaviors, and surge pricing generation. In specifics:

  - The proposed demand and supply generation model can take into account potential spatial, temporal, and modal correlations, as well as the count nature of ridership. Moreover, the model considers impacts of both endogenous (e.g. dynamic pricing) and exogenous (e.g. socioeconomic and land use) variables.

  - The proposed ATS driver partners' platform-exiting model addresses spatial heterogeneity and autocorrelation and clarifies the driven factors for exiting behaviors.

  - It is one of the first studies, to my knowledge, that machine learning approaches have been applied to dynamic pricing generation. As well as

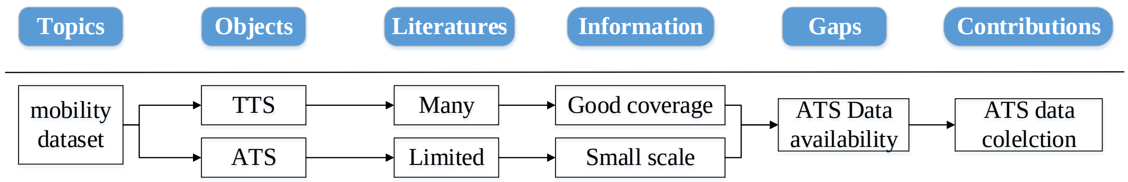


Fig. 1.4.: Topic summary of mobility dataset

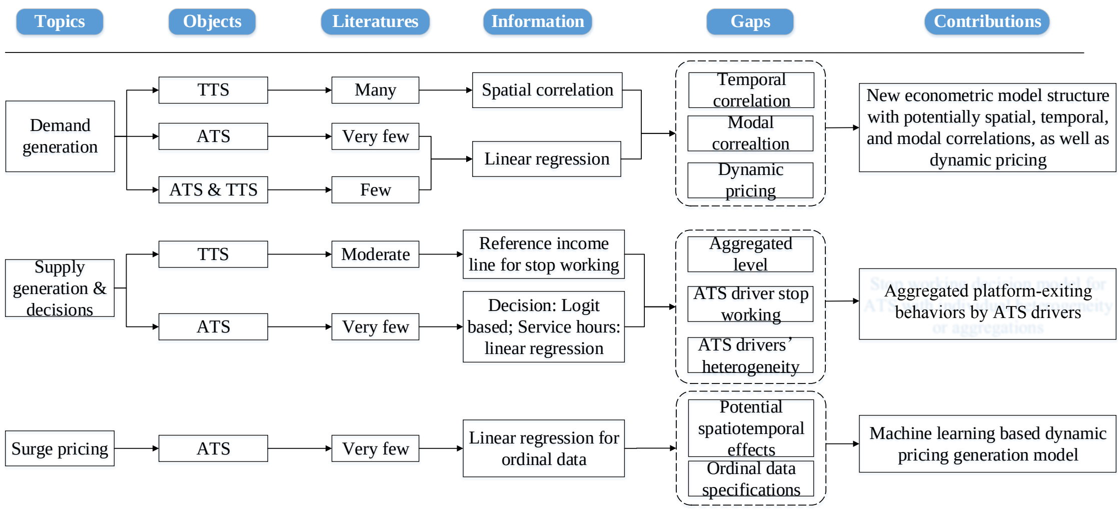


Fig. 1.5.: Topic summary of strategic behaviors

variables in local areas, we also investigate the spatial and temporal heterogeneity while generating dynamic pricing for the whole market.

- **A unified framework of monopolistic competition taxi system modeling**, as shown in Figure 1.6. It provides a holistic understanding of monopolistic competition taxi market dynamics by integrating stochastic processes of passenger-vehicle matching with taxi movements across heterogeneous aggregated regions. In other words, the proposed modeling structure describes not only passenger-vehicle processes within the taxi system, but also interactions with urban road congestion. The new approach is efficient in providing system performance information with many stochastic details, but requires much less

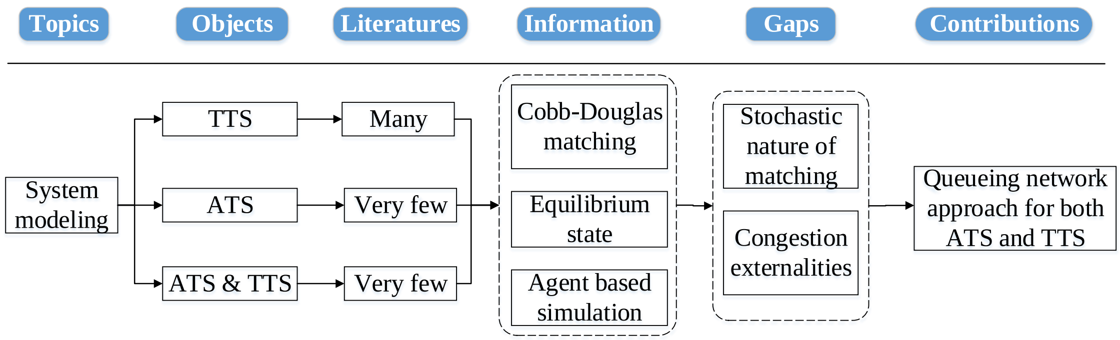


Fig. 1.6.: Topic summary of system modeling & controls

computational effort. More importantly, the system modeling structure, as well as the assumptions, are empirically validated with the availability of large-scale mobility datasets.

1.5 Organization of the Dissertation

This dissertation is organized as follows. Chapter 2 discusses the existing studies on both ATS and TTS, including an overall overview and separate summary of findings for every problem of interest. Chapter 3 presents the data preparation and preprocessing, in particular the acquisition method for ATS with the case of Uber and aggregation scale determinations. Chapter 4, 5, and 6 explore the strategic behaviors of all stakeholders. Specifically, Chapter 4 formulates demand and supply generation under the dynamic pricing. Chapter 5 moves on to aggregated platform-exiting behaviors by ATS drivers. Chapter 6 develops dynamic pricing generation model for discovering relationships among demand, supply, and charges. Furthermore, chapter 7 proposes a holistic model with passenger-vehicle matching and congestion externalities, which allows us to quantitatively measure taxi system performance over spatiotemporal and modal scales. Finally, chapter 8 concludes the dissertation and figure out the future research directions.

2. LITERATURE REVIEW

2.1 Introduction

Taxis play an important role in urban transportation system, not only providing convenient door-to-door mobility services but also supplementing to urban transit system. However, the taxi services are apparently different from urban transit system, by way of decentralized platforms without fixed scheduling and routing. The platform is highly depending upon individual drivers' customer searching strategies, and thus is more likely to be inefficient (e.g. excessive supply at hot spots but shortage at nearby regions) considering drivers' selfishness. Recent technique advances are reshaping the taxi system and enhancing taxi system efficiency with centralized two-sided platforms, as more transportation network companies (TNCs) (e.g. Uber, Lyft, and DiDi) are joining the industry. Regardless of decentralized and centralized platforms, the key components in the taxi system are not changing, consisting of demand and supply generation, supply and pricing regulations, and driver-passenger matching, as shown in Figure 1.3. Understanding these key components with mathematical forms and quantitative analyses is vital to measure the system performance, develop control and management strategies, and plan sustainable urban transportation system.

Analyzing around 850 related articles[1] provides us a preliminary vision on the literature evolution of the field. The field of taxi system modeling is increasing greatly in the last decade, in particular in recent several years, and likely keeps expanding in the next few years. First, Figure 2.1(a) indicates the continuous growth in the number of articles from the year of 2007 (except for the years of 2009 and 2011). Nowadays, there are around 150 articles discussing taxi system modeling each year.

[1]Articles explore taxi demand, supply, driver-passenger matching, system modeling, simulation, and optimization, published from Jan. 1, 2002 to Jun. 30, 2018, and indexed by the Scopus database

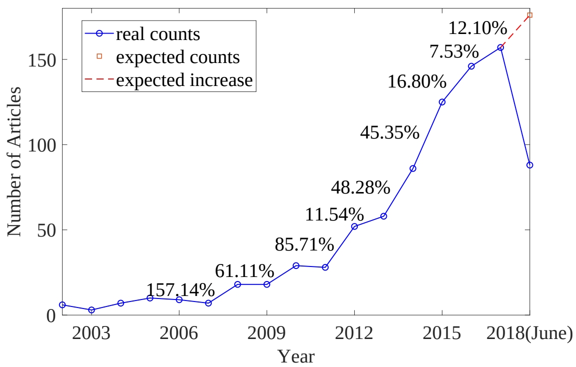
Second, Figure 2.1(b) presents the similar growth pattern in the number of authors per year. Although most articles are contributed by new authors[2], we can still track the increasing portion of active[3] and returning authors[4]. The field is not only attracting new authors, but also producing more continuous works by one unique author. Third, Figure 2.1(c) explores the number of articles by each author and their contribution rate. The authors with more related articles are more likely to have articles in the last three years (red dash line: 50% of articles in last three years), as well as higher contribution rate (blue dash line: average contribution rate across 17 years). Last, Figure 2.1(d) shows the article acceptance by journals. The number of articles published in the last three years are significantly greater than those before 2015. For example, 87% and 83% of related articles are published in Transportation Research Part C (Transp. Res. Part C) and IEEE Transactions on Intelligent Transportation Systems (IEEE Trans. ITS) after 2015, respectively.

In opposition to the accelerating expansion of the field, there are almost no comprehensive survey articles on taxi system modeling. In particular, we have not seen any summaries on challenges and opportunities in the field, due to the rising of big data and centralized platforms. One of the first few survey articles is from Salanova et al. [52], who mainly summarize aggregated models describing relationship among key performance (e.g. vacancy rate, demand, waiting time, and revenue) and equilibrium models for regulation policies. Later, Salanova et al. simply present the generalized mathematical forms for demand, supply, cost, and waiting time, but do not have in-depth discussions and comparisons [53]. The both survey articles are mainly about traditional street-hailing taxicabs with decentralized platforms. Meanwhile, there are another two survey articles regarding ridesharing (one rising application for taxicabs and pilot programs by TNCs). Chan and Shaheen introduce the various forms of ridesharing and present the development in North America [54]. Furuhata et al. also clarify the classification of ridesharing system and, more importantly, point out the

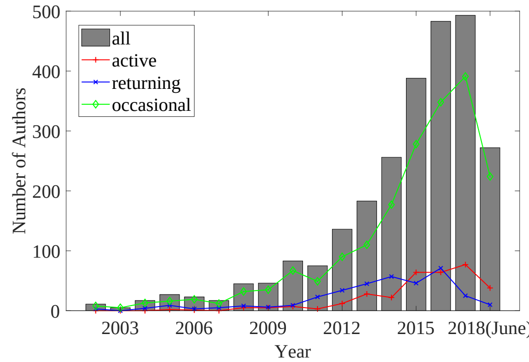
[2]having their first publication on taxi system modeling between the year of 2002 and 2018

[3]having at least one article in the year and at least another one in the last two years

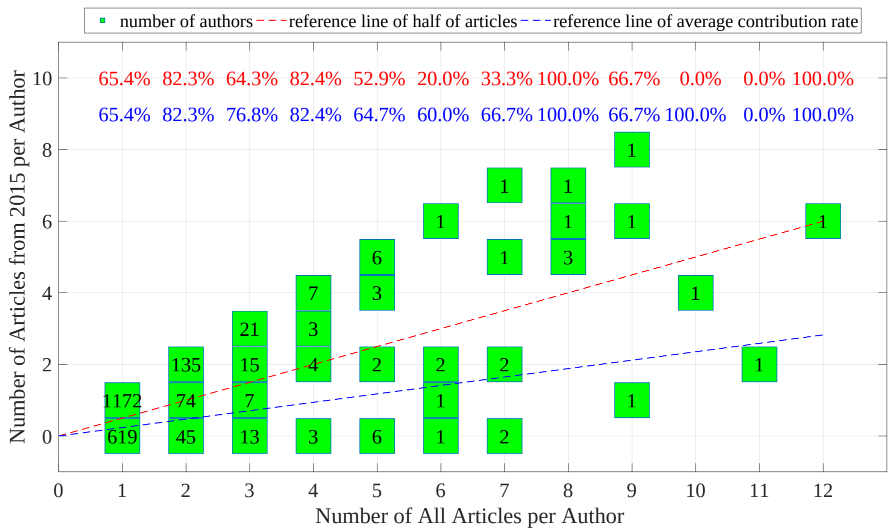
[4]having at least one article in the year and at least another one in more than two years ago



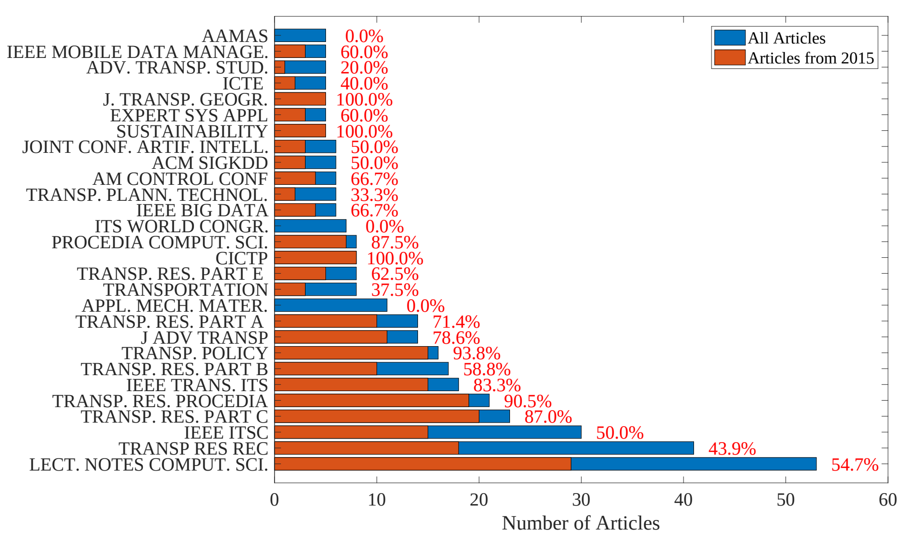
(a) Number of articles



(b) Number of authors



(c) Articles by authors (percentages refer to how many authors are above the reference line)



(d) Articles by publication source

Fig. 2.1.: Bibliographic analyses for publications and authors

challenges from real-time ridesharing problems, for instance, system design, pricing, high-dimensional matching, and social concerns [55]. The both ridesharing survey articles are beyond taxi pooling and involve a broader range of carpool, vanpool, and coworker carpool. In addition, it is apparent that the four survey articles are published before 2013 when the big growth in the field just begins. The survey articles are impossible to cover recent changes and shifts in research methods.

The objective of this survey article is to not only summarize the current methodologies for demand, supply, matching, system modeling and optimization, but also explore challenges and opportunities in the filed as the rise of big data and decentralized app-based taxi services. Towards this goal, we survey the related articles from the year of 2012 to 2018. Except for basic bibliographic analyses shown in the second paragraph, an extensive exploration over topic identification and evolution is implemented with complex network analysis. To the best of our knowledge, there are no any prior such work, tracking topic changes over temporal scale. Then, considering the distinct groups/sides in the taxi system, we review the fundamental methods for their behaviors. Last, starting from the driver-passenger matching, we review the system modeling structures, as well as optimization problems.

The remaining sections are organized as follows: Section 2.2 summarizes the identified topics each year and their year-to-year evolution; Section 2.3 reviews methods for stakeholders' behaviors, including demand generation, driver's customer searching, and other rising problems mainly in centralized app-based platforms; Section 2.4 exhibits the comprehensive modeling structure combining various fundamental components, and points out challenges while understanding taxi system; Section 2.5 concludes the study and indicates extensions.

2.2 Overview of The Field

2.2.1 Topic Identification

The application of complex network for bibliographic database is an elegant but rising solution. Suppose there are n articles $AR = \{ar_1, ar_2, \cdots, ar_n\}$ with m unique defined keywords $KEY = \{w_1, w_2, \cdots, w_m\}$. We can easily derive a bipartite graph $G = \{AR, KEY, E\}$, where E is a matrix of $n \times m$ and consists of $e_{ij} = 1$ if an article ar_i has a keyword w_j. Then, we can obtain another network $KC = \{KEY, B\}$, called keyword co-occurrence network, through $B = E^T E$. In the new network, each node indicates an unique keyword and two nodes w_i, w_j are connected by one link if matrix element $b_{ij} > 0$. Moreover, the magnitude of b_{ij} measures how many times the both keywords w_i, w_j are included in one same article. Given the definitions of keyword co-occurrence network KC, we can create KC every year from 2002 to 2018 (June), as shown in Figure 2.2. Beyond the individual node or link relations over KC, we are more interested in the clusters of nodes (i.e. keywords) that can be interpreted as research themes or topics. This problem can be explored through finding communities over KC network. In the complex networks, community is a group of nodes that can be easily grouped and densely connected internally. Each community over KC is one research topic represented by a set of densely connected keywords. There are multiple methods for community finding. Here, we mainly employ the simple centers algorithm that is a simple but well-known algorithm in the context of co-word analysis [56]. Furthermore, we introduce two metrics of Callon's centrality and density to measure the importance of research topics [56]. The former one is mainly designed for strength of external ties and measures the importance of a topic in the development of the entire research field. In contrast, the latter one measures the strength of internal ties. Based on the mean values for Callon's centrality and density, we can further classify research topics into four categories:

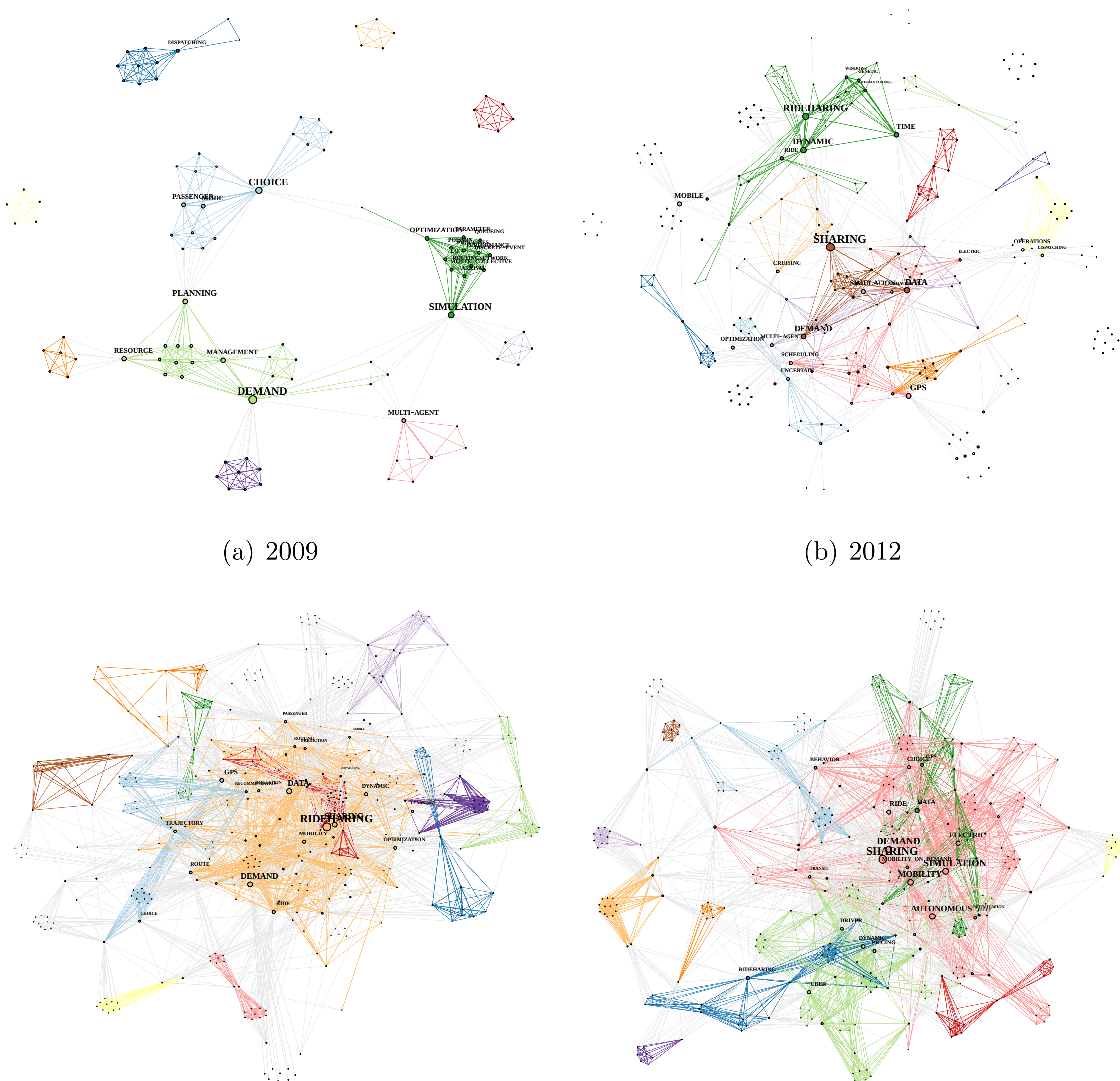
Quadrant I: Topics with larger Callon's centrality and density are both well developed and important for shaping the research field. In addition, topics in this category are externally related to others.

Quadrant II: Topics with smaller Callon's centrality but larger density are specialties in the whole research field. Since these topics are well developed but have rare external ties to others.

Quadrant III: Topics with smaller Callon's centrality and density are weakly developed and less connected to others. They may indicate either emerging or disappearing topics.

Quadrant IV: Topics with large Callon's centrality but smaller density are important for the whole research field but lack in-depth discussion.

Table 2.1 and 2.2 present the topics, as well as their importance, identified from our bibliographic dataset. First, in correspondence with the increasing number of articles, the number of topics in each year is also increasing, from around 5 topics at the beginning to around 20 topics in recent years. Second, the column of Quadrant I (also core topics in the research field) apparently has two phase transitions around 2008 and 2014. From around 2008, topics in the research field become more diverse, covering not only dispatching and regulation but also demand and supply, as well as their interactions. From around 2014, due to the rise of TNCs, autonomous/electric vehicles, and smart computing techniques, the research field explores more problems relating to those new techniques and extends ridesharing/dispatching problem to a real-time or dynamic level. Third, the column of quadrant II is mainly dominated by topics of 'mode choice' or 'demand'. This indicates that those topics are well developed but not tightly connected to the core topics. In other words, the dispatching or taxi system optimization problems have not included demand dynamics as one important concern. Last, from the columns of quadrant III and IV, we can also derive changes in rising techniques, from empty vehicles in the years of 2009 to 2012, real-time large-scale decisions in 2012 and 2013, ridesharing and shareability network



(a) 2009 (b) 2012

(c) 2015 (d) 2018

Fig. 2.2.: Keyword co-occurrence network in four typical years

in 2014 and 2015, stochastic demand and rationality of labor supply in 2016, and route scheduling in 2018.

2.2.2 Topic Evolution

Except for implicit explorations on each year's topics, we derive a topic evolution network. Let $T^t = \{u_1, u_2, \cdots, u_n\}$ be the set of identified topics in the year of t, $T^(t+1) = \{v_1, v_2, \cdots, v_n\}$ be the set of identified topics in the year of $t+1$, and each identified topic may include multiple keywords (for instance, u_1 consists of three keywords w_1,w_2, and w_5). It is said that there is topic evolution from topic u_1 to topic v_1 if there are duplicated keywords. Therefore, we can have a new complex network $G1 = \{T, Evo\}$ with every identified topics as nodes T and evolution relationship as links Evo. Moreover, this complex network can be extended to a weighted one with two types of links through counting duplicated keywords and introducing inclusion index $Incl$ (as equation 2.1). Once there are duplicates in the top 5 keywords of both topics, a solid line connects the both topics. Otherwise, a dotted line is applied. The preliminary plot of topic evolution network for our bibliographic dataset is shown in Figure 2.3.

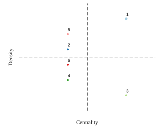
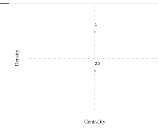
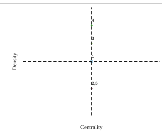
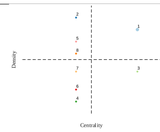
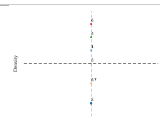
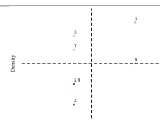
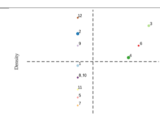
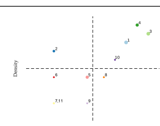
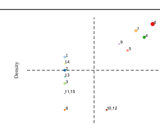
$$Incl = \frac{\#(u_1 \cap v_1)}{min(\#u_1, \#v_1)} \tag{2.1}$$

where, # indicates the number of unique keywords in the identified topic.

Same as community finding while topic identifications, the simple centers algorithm is performed over topic evolution network and explores clusters of topics across multiple years. Overall, we can obtain 7 major clusters, each of which has significant year-to-year topic evolutions, also shown in Figure 2.3. Note that each square marker and number inside indicates one topic in that year and corresponding top keywords are summarized in Table 2.1 and 2.2.

Cluster 5 is mainly from the year of 2002 to 2011 and has most works done before the year of 2010. There are two main research lines within this cluster,

Table 2.1.: Topic Status from the year of 2002 to 2010

Year	Distribution	Quadrant I	Quadrant II	Quadrant III	Quadrant IV
2002		1. Dial-A-Ride	2.Queue; 5. Labor Supply	4. competition, equilibrium; 6. commute modes	3. simulation
2003		1. location based, environment	[no topics]	2. queue; 3. regulation	[no topics]
2004		1. dispatching, dynamic; 3.on-line, scheduling; 4. policy, pricing	[no topics]	2. hierarchical, reinforcement; 5. assignment, fleet	[no topics]
2005		1. ridesharing, efficiency	2. dynamic, flow; 5. congestion, equilibrium; 8. micro searching	4. pooling, dispatching; 6. labor supply; 7.tipping	3. personal rapid transit
2006		1. dispatching; 3. hierarchical, reinforcement; 4. queue, traffic state; 6. ridesharing strategy	[no topics]	5. strategy, graph; 7. monopolistic regulation	[no topics]
2007		1. ridesharing	3. mode choice; 7. adoption, GPS, dispatching	4. automated, dispatching; 6. dynamic pooling	5. sensor network
2008		3. user, mode choice; 4. mobile, acceptance; 6. mode, tourists	2. dynamic ridesharing; 9. pricing; 12. user sense making; new technology	1. simulation; 5. regulation; 7. hub; 8. labor supply; 10. crusing route; 11. allocation	[no topics]
2009		1. mode choice, airport; 3. demand, supply, game theory; 4. simulation; 10. willingness to pay, subsidy	2. dispatching, mobile, adoption	5. multi-agent; 6. bilateral search; 7. matching; 9. on-demand, policy; 11. driver incentives	8. autonomous, utilization
2010		4. dynamic ridesharing; 5. sharing, moblie; 6. simulation, dispatching; 7. data, recommendation; 9. driver, customer, game theory	1. assignment; 2. regulation, welfare; 14. empty trip	3. adoption; 8. pricing; 11. labor supply; 13. bilateral search; 15. user trust, social network	10. vacant taxi; 12. hot spots, user

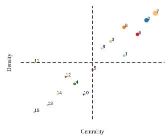
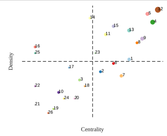
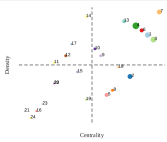
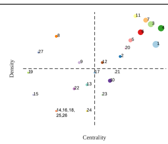
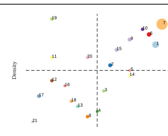
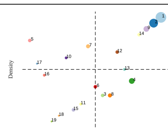
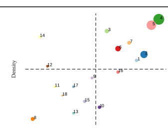
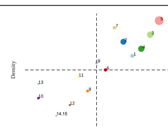
Quadrant I: topics are both well developed and important for framing the field of taxi system modeling;

Quadrant II: topics are well developed but not important for leading the field;

Quadrant III: topics are either emerging or disappearing;

Quadrant IV: topics are important for framing taxi system modeling but not well developed.

Table 2.2.: Topic Status from the year of 2011 to 2018

Year	Distribution	Quadrant I	Quadrant II	Quadrant III	Quadrant IV
2011		1. spatial-temporal, search; 2. crusing, passenger, simulation 3. agent, uncertainty 6. mobile, challenges; 7. ridesharing, dynamic, optimization; 8. airport; 9. mobility, fleet	11. dial-a-ride	4. recommendation; 10. ridesharing, performance; 12. search frictions, equilibrium; 13. flat rate; 14. empty, redistribution; 15. ridesharing	5. regulation, pricing
2012		1. agent, uncertainty; 4. dynamic ridesharing; 5. GPS, electric, scheduling; 8. intelligent decisions; 9. multi agent simulation; 11. dispatching, redistribution; 12. sharing, demand; 13. mobile, on-demand; 15. optimization, responsive; 23. commute share	14. clustering, recommendation; 16. mobility-on-demand; 25. pooling	3. ridesharing, mobility; 10. routing; 17. taxi volume, phone call; 18. fleet, charging; 19. guidance; 20. reinforcement; 21. carsharing; 22. rideshare, reliability; 24. airport pickup; 26. taxi dispatching	2. real time, large-scale, decisions; 6. vacant, passenger wait; 7. cruising, impacts
2013		1. simulation, sharing; 3. passenger, recommendation; 4. dynamic ridesharing; 6. airport, choice; 7. GPS, demand, prediction; 9. supply; 10. fleet, charging; 13. smart application	11. sustainable; 12. simulation, fare; 14. mode choice; 17. scheduling	15. reinforcement, sequential; 16. matching; 19. location prediction; 20. personal rapid transit; 21. empty; 22. Bangkok; 23. large-scale, pool; 24. interaction-rich transport	2. mobility-on-demand, optimization; 5. intelligent carpooling; 8. mobile, allocation; 18. tourist, seasonal
2014		1. dynamic ridesharing; 2.spatial, pickup, prediction; 3.electric, pricing, optimization; 4. data, simulation, demand; 5. search friction; 6. mobile support; 7. driver, mode choice; 11.management, distributed; 12.cruising; 20.demand	8.personal transit; 9.route; 27.taxi assessment	13.app adoption; 14.ridesharing, optimization; 15.share-a-ride, freight; 16.e-hail, protocol; 18. electric taxi; 19.location based service; 22.for-profit, strategy; 24.demand-supply equlibrium; 25.customer search; 26.unoccupied, hubs	10.on-demand; 17.shareability, network; 21.dynamci ridesharing; 23.automous fleet
2015		1. GPS, recommendation, choice; 2. path; 6. mobility-on-demand, control; 7. dynamic ridesharing, data, demand, optimization; 9. vacant, sequentail behavior; 10. regulation, environment; 15. autonomous	11. game, bargain; 19. stochastic demand; 20. commute mode	8. pooling, station; 12. online booking; 13. responsive transit; 16. carpooling, sustainable; 17. time window, delivery; 18. multi-agent, artificial; 21. schedule, delay	3. bayesian, UBER; 4. ridematching application; 5. campus pooling, choice; 14. on-deman, smart
2016		1. dynamic ridesharing, optimization, recommendation; 2. sharing, behavior, choice; 9. on-demand, UBER, pricing; 12. meeting points; 14. OD infer	5. regression, decision, airport; 7. mobile booking; 10. mobility, OD; 17. opportunistic riders	11. arrival, non-homogeneous; 15. efficiency, graph; 16. earnings; 18. sharing, participation, motivation; 19. trip request prediction	3. disruptive innovations; 4.intelligent matching; 6. queue, decision; 8. personalized recommendation; 13. larbor supply, rationality
2017		1. data-driven, optimization, approximation; 2. online, choice, UBER; 3. game, duopoly; 4. demand, dynamic ridesharing, optimization; 5. passenger, driver, prediction; 6. surge pricing, equilibrium; 7. arrivals, markovian	12. acessibility; 14. MFD, network	8. regulation; 9. logit, mode choice; 11. customer-search; 13. dial-a-ride; 15. route; 17. empty duration; 18. lifesytle, trips	10. electric, sustainable; 16. consumer preference
2018(June)		1. users, search efficiecny; 2. ridesharing, online; 3. dynamic pricing, UBER, ride; 4. GPS, dispatching, optimization; 5. sharing, autonomous, demand, simulation; 7. mobile acceptance	9. e-hail adoption	8. congestion; 10. carpooling, attitudes; 11. neural network, predict; 12. feeder, competition; 13. game theoretic; 14. safety, usage; 15. customer satisfication	6. route, scheduling

Quadrant I: topics are both well developed and important for framing the field of taxi system modeling;
Quadrant II: topics are well developed but not important for leading the field;
Quadrant III: topics are either emerging or disappearing;
Quadrant IV: topics are important for framing taxi system modeling but not well developed.

including the dispatching, ridesharing, and regulation for the first line and mode choice for another line.

Cluster 21 is mainly from the year of 2008 to 2013. This cluster address the problem of pricing and fleet regulation policies with game theory and search frictions. Several minor topics are also mentioned in this cluster, including carpooling, demand responsive services, and multi-agent simulations.

Cluster 30 is mainly from the year of 2010 to 2014. Although this is a relatively small cluster, it solves a critical problem in the taxi operation, about empty vehicle movement and redistributions.

Cluster 27 is mainly from the year of 2010 to 2015 and has most works done before the year of 2014. One apparent characteristic of the cluster is data-driven analyses, varying from location recommendation, large-scale and real-time decisions, demand prediction, to matching.

Cluster 26 is mainly from the year of 2010 to 2016. This cluster mainly measures the performance of taxi system and quantify the benefits from inclusions of new techniques, such as electric and autonomous vehicles. Within the studies, the dynamics of passengers and drivers (in particular empty vehicle drivers) are also discussed.

Cluster 33 is mainly from the year of 2013 to 2018. The major topics in this cluster are still about taxi system control and management, as well as dynamic ridesharing designs and benefits. In addition, we have seen preliminary discussions on app-based taxi service, describing the adoptions and users' preferences.

Cluster 40 is mainly from the year of 2015 to 2018. This cluster is a bit different from other clusters, since it mixes various topics together other than one or two representative topics. In this cluster, we can see not only different types of taxi services and designs (e.g. electric, app-based, on-demand, and autonomous) but also various problems from ridesharing, dynamic pricing, passengers' acceptance, drivers' behaviors, to interactions between drivers and passengers. Moreover, the tight connections among those topics also indicate a new trend in the research field that a unified taxi system modeling framework for quantifying and planning is gradually involving behavioral models for various stakeholders, for instance, drivers, passengers, service providers, and even society system.

Figure 2.4 presents the number of articles in each cluster from the year of 2002 to 2018. The trends are also in correspondence with three stages identified in above analyses. Before 2010, the research field is mainly dominated by cluster 5 of 'regulation, ridesharing, and dispatching', contributing no more than 20 articles per year. Then the research field shifts to the second stage from 2010 to 2013, likely driven by the rise of big data and GPS trajectory of taxicabs. The high-resolution dataset allows us to have in-depth discussions on vehicle movements, in particular empty vehicles in cluster 30, and have a perfect knowledge on spatiotemporal distribution of demand and supply in cluster 27. In addition, this stage also extend studies for taxi system regulation policies. The number of articles doubles to around 50 per year. Last, the research field is experiencing the third stage from 2014, likely driven by the advances of techniques, for instance, e-hailing apps and autonomous/electric vehicles. This stage has not only explosive growth in number of articles per year but also frequent changes while developing topics. At the beginning of this stage, the research field focuses on quantifying the performance and benefits of both traditional and emerging taxi services, mainly in cluster 26 and 33. Then it shifts to cluster 40 and deeply explore behaviors of all stakeholders and unified system modeling frameworks for various types of services.

2.3 Strategic Behaviors of Stakeholders

As illustrated in Figure 1.3, the taxi system in general is a multiple-side platform, consisting of passengers, drivers, and centralized agency or TNCs. Before reviewing

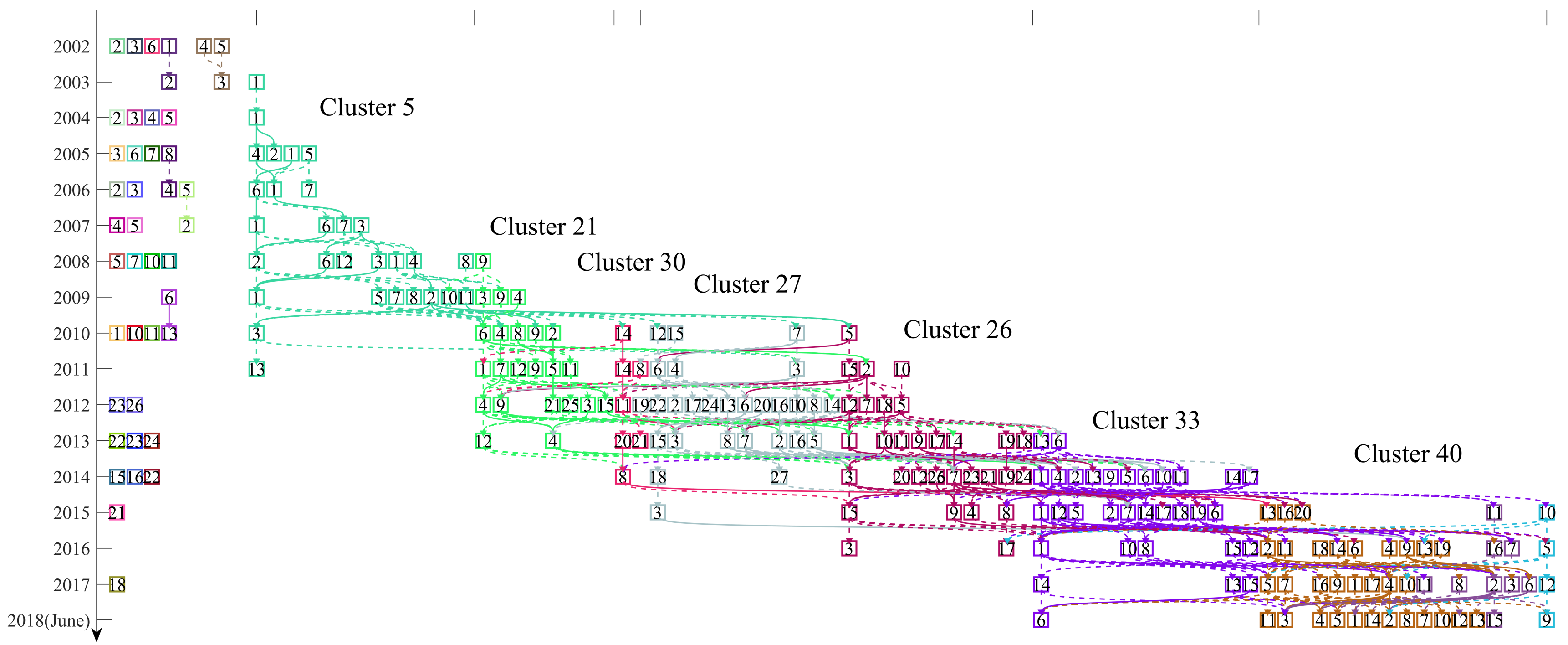


Fig. 2.3.: Topic evolution in the last decade

unified modeling framework based on driver-passenger matching, this section primarily summarizes the current progresses in each side of taxi systems, as well as their interactions. Figure 2.5 illustrates major behaviors of passengers and vehicles. For passengers, we are interested in the demand generation (behavior 3) or where/when the passengers generate. Limited by data availability, most studies are utilizing the real number of pickups as demand. This underestimates the taxi demand, since it excludes those unserved passengers quiting for other transportation modes. Therefore, the term 'demand' in this section indicates taxi pickups (or served passengers) rather than the real demand. On the other hand, the drivers' behaviors are also of interests and contains multiple problems, such as behavior 1 - supply generation (where/when driver partners generate) and 2 - stop working (where and when driver partners stop works), behavior 4 - customer searching behaviors of empty vehicles (on which route, to where, and how long) and behavior 5 - total working hours. Different from TTS with fixed fleet and pricing regulations, ATS generally applies real-time pricing control, called surge or dynamic pricing. Beyond the passengers and drivers, the centralized TNCs can also intervene individual behaviors mainly through adjusting fare rate and fleet size, which are not shown in Figure 2.5. To sum up, this section exhibits the overviews of current understandings over passengers', drivers' and centralized platform owners' behaviors.

2.3.1 Demand

Based on the study period and objects, we can further divide related studies into three groups, including group 1 - long-term demand generation over a relatively longer time period; group 2 - short-term demand generation within a short time interval; and group 3 - OD derivation and mode choices.

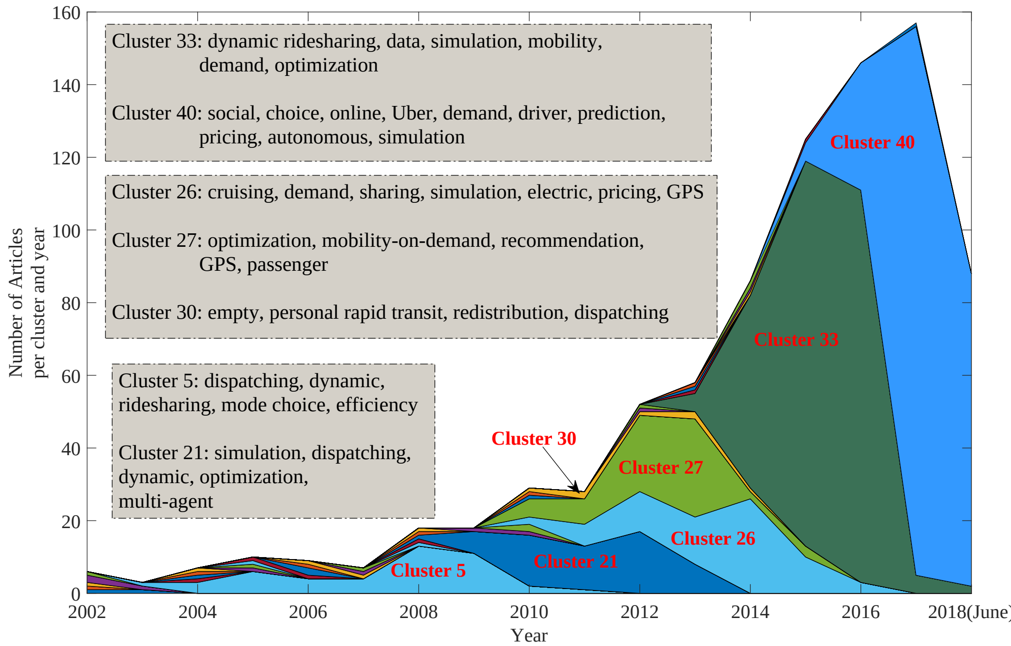


Fig. 2.4.: Number of articles by topic clusters

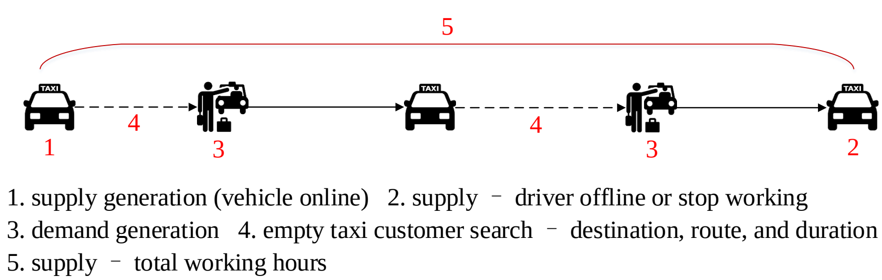


Fig. 2.5.: Illustration of strategic behaviors in the taxi system

Long-term Demand Generation

At the beginning, the taxi demand is modeled mainly with a generalized mathematical form and endogenous variables, for instance, waiting time and trip fares. Schaller measured the price elasticity of -0.22 for trip demand in NYC, using the multiple regression model [57]. Yang et al. inferred daily taxi trips with waiting time, fares, trip travel time, and passenger price index for transportation, as well as total population and average income [12]. In one of their later articles, the demand function is developed as a strictly decreasing and differentiable function of the full trip price and the full trip price can be formulated with expected fares, waiting time, and trip travel time [58]. Guo et al. recently examine ATS rides in the case of Shenzhou Ucar service and related to the dynamic pricing, under the power-law function [51].

As the rise of big data, it becomes easier to obtain not only huge taxi trip records but also more exogenous variables from the society system. There also emerges much more studies on the influencing factors with econometric models. Kamga et al utilized 147 million Yellow Taxi trip records for 10 months in NYC to measure the correlations between traditional taxi ridership and variables such as the time of day, days of the week, and weather [59]. Yang and Gonzales further developed an multivariate regression model for hourly pickups and drop-offs and explored the impacts of population, education, age, income, employment, and public transit accessibility [29]. Lacombe and Morency utilized multivariate regression model for daily pickups and drop-offs in Montreal with variables describing weather, demographic, land use, and transit accessibility [60]. Ge et al. also applied multivariate regression models on taxi ridership in two large cities - NYC and Shanghai - and identified significant impacts of urban land use, resident population, employment, and vehicle ownership [30]. Zhang et al. conducted a zero-inflated negative binomial model to analyze temporal characteristics of NYC Yellow Taxi trips, as well as the impacts of trip distance, urban built environments, and socioeconomics [31]. Frechette et al. employed a generalized regression for the specifications of number of waiting customers by waiting time and time effects [21]. Neoh et al. reviewed 22 existing empirical studies with meta-analysis and summarized the influencing factors, such as employees, partner matching programs, work schedule, and gender [61]. Cohen et al. obtained 50 million individual-level observations on Uber ride-callings and introduced a regression discontinuity framework for relationship among purchase rate, surge pricing, and waiting time [47]. Brodeur and Nield investigated the impacts of rainy hours on taxi, Uber, and Lyft ridership, as well as service competitions, utilizing three separate multivariate regression models for different services [62].

Beyond the econometric analyses for influencing factors, the research field is also thinking about the modeling structures given the facts of significant spatial, temporal, and even modal correlations. Most taxi systems, in particular in larger cities, have unbalanced ride distribution, concentrating in downtown areas and rarely picking up in remote suburban areas. This leads to an interesting spatial autocorrelation that areas with high (low) ridership are also surrounded by areas with high (low) ridership. Qian and Ukkusuri introduced a geographically weighted regression model on taxi ridership that can account for spatial effects. They also verified several additional exogenous variables on ridership of TTS, such as education level, commercial areas, income, road density and subway accessibility [32]. Another similar study by Nam et al. also employed geographically weighted regression to account for spatial effects and specify taxi ridership with variables of transit and city density in Seoul [33]. Zhang et al. adapted a spatial lag model, one of spatial econometric methods, for the ridership of both ATS and TTS daily ridership and identified the impacts of socioeconomic, land use, built environment, and day-of-week [63]. Moreover, Zhang et al. explored the modal correlations in the taxi system coexisting of both ATS and TTS through combing spatial lag models with a simultaneous equation structure. The model performance is better than a separate model for each service, which indicates unobserved correlations between ATS and TTS. Last, as other transit modes in urban areas, taxi ridership also exhibits certain hour-to-hour, day-to-day, and even week-to-week correlations. However, we have not seen any progresses on such problems.

The temporal correlations, along with spatial and modal ones, make the econometric analyses more interesting but also challenging.

Short-term Demand Generation

In contrast to the long-term demand generation, a group of studies also focuses on the short-term taxi demand generation that are the number of passengers in the next immediate time interval and are of interest for real-time system control and management. In general, these studies work on the sequence of taxi demand based on time at one fixed location and propose methods for temporal correlations. The key approach in such studies is the integrations of historical number of pickups to predict the immediate future one. Moreira-Matias et al. combined three time-series forecasting techniques that are weighted time-varying Poisson model, autoregressive integrated moving average, and ensemble learning, to predict short-term taxi ridership at each defined taxi stand [64]. Zhao et al. proposed three predictors, that are Markov, Lempel-Ziv-Welch, and Neural Network, and compared their performance with taxi demand prediction at each city block [65]. Qian et al. extend the temporal scale to spatiotemporal one with Gaussian conditional random field method, considering number of taxi pickups in the previous time intervals at both same location and neighboring locations [66, 67]. Later, Liao et al. introduced two types of deep neural networks and confirmed the advantages in prediction accuracy over other machine learning approaches but highly depending on the deep neural network structures [68]. Moreover, the proposed deep neural networks can include the impacts of external factors, such as weather and holiday, which has not been done in aforementioned short-term studies. Faghih et al. compared two spatial-temporal time-series models, called spatial-temporal autoregressive (STAR) and least absolute shrinkage and selection operator applied on STAR (LASSO-STAR), and verified the better performance by LASSO-STAR for Uber demand. More importantly, this study also presented the improvements on prediction accuracy while considering spatial effects ignored in most short-term forecasting [69]. Xie et al. assumed that ATS demand generation is related to all other transportation modes, hence utilized multiple ridership by TTS, buses, metro, and private vehicles to estimate ridership by ATS. However, there are almost no unified modeling framework for short-term predictions of both ATS and TTS ridership, meanwhile considering both spatial and temporal relations.

OD Derivation and Mode Choice

Additional problems are also explored other than number of pickups or passengers in a fixed time period and location. The first one is unmet passengers estimation. Afian et al. defined an unmet intensity metric with the ratio of taxi demand to excess supply and approximated unmet passengers [70]. However, such problems are hard to measure real passenger demand and subsequently calibrate models. The second one is the passengers' destination or passenger OD matrix. The elements of passenger OD matrix indicate the number of trips from origin to destination. Their estimations are same as the demand generation with endogenous variables, for instance, expected waiting time at origins and trip costs from origin to destination [71]. On the other hand, the probability of passengers traveling from origin to destination may be also estimated, given the potential demand at origin. Almost all studies related to the destination choice probabilities are built on the expected utility method, as well as exponential- or logit- form. Wong et al. proposed an exponential function with trip cost, waiting time at origin, and travel time between origin and destination [72], followed by Grau and Estrada [73]. Yang et al. derived an utility function with corrected trip cost consisting of charges and monetary values for times and computed destination choice probabilities under logit-based modeling structure [17]. Lavieri et al. extend the logit functional form to a quasi-likelihood estimation approach while splitting ridesourcing demand to multiple destinations [74].

The last one is on the mode choice between ATS and TTS or between taxi and other urban transportation modes. Qian and Ukkusuri measured the monetary trip cost with fare, waiting time and travel time, and simulated passengers' choices between ATS and TTS according to minimal costs [26], which is similar to the generalized trip costs and logit-form models by Wang et al. [25] and He and Shen [24]. Wong et al. proposed a hierarchical logit mode choice between normal traffic and taxis, also based on generalized costs of the alternatives [16]. The other efforts introduced more exogenous variables obtained from stated preference survey. Zhang et al. implemented a binary logit model on mode choice between ATS and TTS with socioeconomics, perceptions over cost effectiveness, safety, and convenience, and characteristics of commuting trips [41]. Shaheen et al. completed 503 surveys on casual carpooling users in San Francisco, introduced a logit modeling for mode choice with age and employment status [4]. Dias et al. introduced a Bivariate ordered probit model to formulate the choice behaviors between ride-sourcing and car-sharing services, which can assist in travel demand forecasting of mobility-on-demand services [43]. There is one additional study utilizing machine learning approaches other than discrete choice modeling to understand binary mode choice. Chen et al. modeled the ridesharing choice as a general binary classification problem taking into account features for instance departure time, trip length, fee, weather, and travel time, based on the real example of DIDI in China [48]. However, there is a lack of empirical evidence on the performance of machine learning approaches vs. discrete choice modeling. Moreover, machine learning approaches learn knowledge directly from data and requires high resolution and reliable training datasets and cannot leverage current available human experiences, prior knowledge, or findings on transportation mode choices.

One apparent characteristic of these studies is the logit-based discrete choice modeling structures that are based on expected utility theory. As a classical assumption in decision theory, the expected utility has been utilized for many years and in various fields. It assumes rational behaviors based on maximum utility and equal attitudes on losses and gains and has evidence of validity. However, it is still questioned in recent years. The advanced decision theory that intervenes the psychological and social details, processes differences among heterogeneous people, captures irrational behaviors, and addresses loss-aversion [5] should be favorable and reliable. The field of transportation also begins to apply prospect and regret theory into discrete choice problems but not specialized to taxi system.

Summary

The demand generation, as well as mode and destination choice problems, are critical concerns while understanding taxi system. We review considerable findings on the taxi ridership and empirically identified influencing factors (both endogenous and exogenous). However, these studies are limited by modeling structures and fast-changing taxi markets, as the big data techniques yield better understandings of large-scale taxi system. **First**, as we extend the analyses to a large-scale system, the spatial effects play a vital role in the modeling, regardless of short-term and long-term. In the traditional econometric models used by long-term analyses, spatial autocorrelations violate the underlying assumptions of normality, homoscedasticity, and uncorrelated errors thus require additional spatial processes (e.g. spatial lag, spatial error, and geographically weighted regression). **Second**, the potential temporal correlations from hour (day/week/month) to hour (day/week/month) have not been addressed well in the long-term analyses. **Third**, the rise of ATS has reshaped taxi industry and improved the system variety. While modeling the demand generation, it is inevitable to consider modal correlations in terms of competitions and supplements. However, almost all studies independently explore ridership by services and do not have impacts from other similar services. **Fourth**, due to the complicated interactions at spatial, temporal, and modal scales, we have not reached any agreement on modeling structures and also lack comprehensive comparisons even at one single scale. Taking the spatial autocorrelations as an example, the weight matrix derivations, as well as number of correlated neighborhood, are still under study. **Last but not the**

[5] Loss-aversion is the dislike of losses, compared to gains. In other words, individuals tend to dislike losses much more than they like gains

least, current progresses examine impacts of endogenous or exogenous variables independently and lack comprehensive views on the both. Due to the real-time controls over ATS, both endogenous (e.g. dynamic pricing) and exogenous (e.g. weather and land use) variables are important while specifying ATS and TTS ridership.

2.3.2 Supply

Based on drivers' decisions and vehicle status (occupied or empty), we can category current progresses into two groups, including group 1 - service availability (in terms of number of vehicles and working hours [behavior 5 in Figure 2.5]) and enter/exit decisions (behavior 1 and 2 in Figure 2.5); and group 2 - empty vehicle movements between one drop-off and next pickup (behavior 4 in Figure 2.5).

Fleet and Drivers' Enters/Exits

The fleet indicates number of registered taxicabs in TTS and number of active drivers (generally have more than 4 rides in one month) in ATS. Schaller employed multivariate regression model for impacts of service availability on the total number of taxicabs and quantify the elasticity as 1.0 in NYC [57]. In another similar study, he summarized number of cabs in 118 US cities and identified three main influencing factors, including commuters by subway, households without vehicles, and airport trips [34]. Kamga et al. checked the statistical correlations between taxis on duty and weather [59]. Yang et al. introduced a Wavelet Neural Network for predicting number of taxicabs, which can address nonlinearity relationships with explanatory variables [75]. Hall and Krueger implemented a comprehensive analysis of Uber's labor market in major US cities and statistically measured correlations between supply and socioeconomic variables [6]. Qian and Ukkusuri relates the number of new online ATS drivers to a set of variables including surge pricing and current trip requests [26]. Similar as short-term demand generation, the number of entering drivers are also estimated based on time-series approaches [76]. In addition, several efforts are also on the total working hours or mileages. Flores-Guri proposed a linear function for log transformation of supply (in terms of working hours) and a set of explanatory variables, such as price, demand, fleet size, and minimum wage [77]. Chen proposed a generalized regression model for service duration and found that dynamic pricing can significantly increase supply [38]. Schaller measured the relationship between fare rate and taxi service availability (in terms of empty travel mileages) by a multivariate regression model [57]. Overall, the fleet size and total working hours are not frequently mentioned in both TTS and ATS. The analyses are mainly based on simple statistical and econometric models. Compared to demand generation, we may require more deep visions on supply generation, in particular by ATS driver partners.

In contrast, the stop-working behaviors of taxicab drivers are extensively discussed from both theoretical and empirical perspectives. Farber developed discrete choice model and showed that taxicab drivers' stop-working behaviors are primarily related to cumulative daily working hours other than daily income [78]. Frechette explored the utility of taxicab drivers for shift behaviors (start and stop) and derived conditional probability based on revenues and system uncertainties [21]. Later, the literature mainly focused on reference level of income and verified empirically. Farber incorporated reference-dependent preferences into proposed modeling structure and concluded that there may exist a reference level of income on a given day but it varies substantially day to day for a particular driver. Moreover, it is found that most drivers may end service before reaching their reference levels of income [79]. Agarwal et al. also confirmed the existence of a reference level of income and additionally observed rational drivers' behaviors that can be described with neoclassical economic model [80]. Martin modeled the behaviors by taxicab drivers under prospect theory and indicated that the behaviors are consistent with prospect theoretic (S-shaped) reference dependence as opposed to more extensively examined loss aversion model of reference dependence. Furthermore, prospect theory are proved to have explanatory power for such problems [81]. Nar et al. proposed time-dependent reference point and corresponding learning methods based on hidden Markov model and expectation-maximization algorithm, while applying prospect theory for taxicab drivers' stop-working behaviors [82]. Regarding the ATS drivers, Chen proposed a logit-based discrete choice model for Uber driver partners' offline decisions. The results shed light on the significant impacts of dynamic pricing, given the facts that driver partners work more at times when expected earnings are high (e.g. high surge times) [38]. The conclusion is similar as another study by Horton, who proved that Uber drivers are sensitive (highly elastic) to sudden fare changes, likely because of minimal barriers to entry [83]. However, we have not seen any solid evidence on the reference level of income for ATS driver partners. This requires additional studies for ATS driver partners other than similar existing ones for TTS drivers, since ATS driver partners consist of both full-time and part-time (a large portion) drivers and behave under dynamic pricing (at least 20% increase over standard fare rate).

Empty Vehicle Movements

The empty taxicabs are irremovable from taxi system, even the emerging TNCs. This is because of dislocations at both temporal and spatial scales while generating demand and empty vehicles. After one passenger drop-off, the taxicab or TNC vehicle (no consideration of shared rides) becomes empty and begins to cruise or circulate on streets to find next passengers. Therefore, investigating empty vehicle movements and enhancing vehicle utilization rate play important roles in taxi system operation. Originally, such problems are limited to locations of interest (e.g. taxi stands) or zones and explore whether drivers drive there for next passengers with simple logit based method. Wong et al. proposed a sequential logit approach to model bi-level decisions at taxi stands, including whether they will travel to the taxi stands and whether they will wait there for customers after arrivals. The study was based on a stated preference survey and identified the significant impacts of distance, drivers' preferences, and congestion [36]. Szeto et al. implemented a sequential binary logistic regression model for the effectiveness of off-road taxi calling installation [84]. Sirisoma et al. implemented a stated preference survey of 400 taxi drivers and developed a multinomial logit model to simulate taxi drivers' customer searching behaviors. They also quantified impacts of endogenous variables, for instance, waiting time, trip time, travel distance, and fare [85]. These findings were similar with another study but on GPS data from 460 taxicabs [86]. Szeto et al developed a time-dependent logit model to model vacant taxi movements and indicated significant impacts of demand, earned fare, and time of day, using GPS data from about 460 taxicabs [35]. Zheng et al also assumed that there are no any pickups along the way to final destination and made decision based on reward functions and logit-based method [87]. In addition to these independent studies solely on empty vehicle movements, such problems are also one important component while modeling taxi system and formulated with logit-based methods [16, 17, 20, 24, 71, 72].

Later, such problems are extended to a zonal level choices rather than location level. More importantly, the following studies assumed that the passenger finding process is not only related to origin and destination but also influenced by a sequence of zone-to-zone choices along the searching routes. Wong et al initially extended the simple logit form to sequential processes and stated that empty vehicle movements were significantly influenced by probability of successfully picking up a passenger along their routes [88, 89]. Moreover, Wong et al. combined a zonal choice model with the sequential logit models and proposed a two-stage approach for taxi passenger finding [90]. Tang et al. proposed a similar two-layer modeling structure but introduced additional routing choices among predefined available routes, rather than given searching routes [37]. Long et al. extended Wong et al's studies and assumed that each taxi driver travels towards an adjacent neighborhood to maximize expected rate of return, a value computed from cumulative profits, empty duration, and expected travel time of next ride [91].

Almost all studies are about TTS drivers' behaviors, who are generally search customers based on their own knowledge. However, the introduction of smart-phone based apps makes ATS drivers relatively more informative to real time demand distribution and exact passengers' origins. The passenger finding process by ATS are expected to be a bit different from those by TTS, once those information is aggregated across various temporal and spatial dimensions [92]. Moreover, we have not seen many inclusions of sequential passenger finding models into taxi system modeling, although the sequential processes empirically outperform simple logit-based models. The challenges remains as to how some of the more behaviorally appealing models are incorporated into a large scale and complex taxi system modeling. Last, the current progresses are still suffering from the underlying assumption of expected utility theory and incapability of addressing individual heterogeneity, regardless of logit-based models and sequential processes.

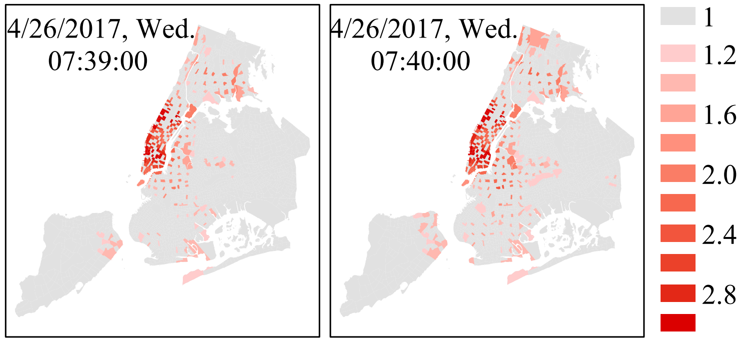
2.3.3 Rising Concerns with ATS

Dynamic Pricing

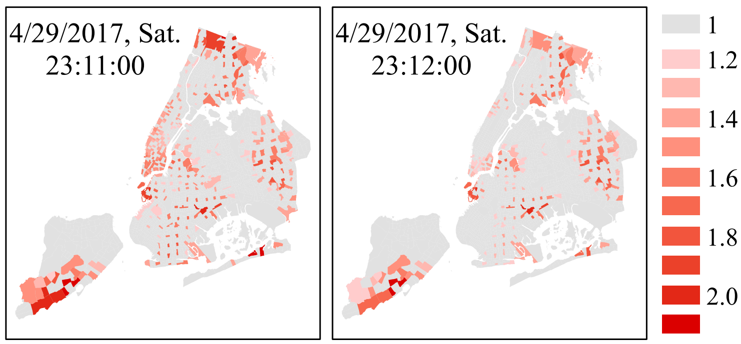
The traditional taxi market has a widely adopted pricing structure, consisting of initial fare, time- and distance-based charges, and additional surcharge fees. The fare rate is fixed for all three parts and may be valid for multiple years. Whenever and wherever you hail a taxicab, the fare rates are same or have slight increase in surcharge fees during peak hours. Except for fare rate settings based on the above pricing structure, we have also seen multiple efforts on pricing structures for TTS. Yang et al. examined a nonlinear pricing structure with only initial fare and distance based charges, under which distant (short) trips have relatively lower (higher) cost, compared to linear pricing structures [93]. Gallo started from the equity and proposed an OD-based fare structures to address asymmetry distribution in taxi services and reduce fares on OD pairs poorly served [94]. Zeng and Oren introduced discounts or surges depending on number of passengers in one ride (or vehicle occupancy) and recommended a lower fare for group rides [95]. Qian and Ukkusuri developed a time-of-day pricing structure, under which fare rates vary by hours and optimal rates are set up based on maximum market revenue [96]. However, TTS is still favorable with long-standing fixed linear pricing structure and has not implemented any pilot program on above innovative pricing structures.

On the contrary, ATS employs a dynamic pricing mechanism with both a standard fare rate as TTS and a dynamic multiplier varying from 1.2 to more than 4.0. The dynamic multiplier is designed during supply shortage and updates frequently (every 4 to 5 minutes), as shown in Figure 2.6. ATS platforms track real-time distribution of demand and supply and balance them in the way of decreasing ride requests but attracting additional driver partners with a higher fare rate. Theoretically speaking, dynamic pricing is important for efficient mobility markets and has positive impacts on welfare. However, the debate over surge pricing has polarized opinions by the general public, due to transparency of surge pricing generation and public misunderstandings. Suranovic began with the public misunderstandings on surge pricing, analyzed reasons such as market imperfection, and suggested that demonstration effects [6] are the most effective way of surge pricing implementation [97]. Very limited studies have obtained empirical dataset on dynamic pricing and explore underlying mechanisms. Hall explored Uber's surge algorithm at two hot-spots in NYC and found that the surge algorithm filtered demand and encouraged supply such that a ride was almost always fewer than 5 minutes away, regardless of demand conditions [98]. Guo et al. defined a ratio of accepted orders and figured out the relationship with dynamic pricing by curve fitting techniques [39]. Later, they also proposed two predictors based on Markov chain and neural networks to capture dynamic pricing multiplier in immediate future based on historical multipliers [99]. Chen et al. analyzed the changes in dynamic pricing multiplier over times and empirically verified the correlations between supply, demand, estimated waiting time, and surge pricing multiplier. Moreover, the linear regression model was developed to predict surge multiplier in the next 5 minutes, but did not perform well given a lower goodness-to-fit R^2 value less than 0.5 [40]. Overall, it is apparent that existing studies are still limited by not only

[6] Demonstration effects are effects on individual behaviors caused by observation of actions by others and their consequences.



(a) One typical Wednesday



(b) One typical Saturday

Fig. 2.6.: Dynamic pricing multipliers of UberX in New York City

(Uber data are collected based on 470 data collection points across the city, details in [100])

data availability but also a well-developed modeling structure. We are still not clear about how the surge pricing interacts with demand, supply, and even surge pricing itself in previous time interval and neighboring regions.

Differentiated Operation Strategies

Apparently, the rising ATS provides not only diverse mobility services (from shared rides, economy, to premium) but also different operation strategies at spatial, temporal, and even individual level. At spatial scale, regions with frequent supply shortage (e.g. Manhattan and city borders in Figure 2.6) are more likely to have higher surge pricing multiplier, regardless of how long passengers travel and how many passengers share one single ride. At temporal scale, even one same region also have significantly different surge pricing multiplier across hours and even minute-to-minute, as shown in Figure 2.6. The last interesting but also likely ignored point is at individual driver partner. For newly registered driver partners, most platforms will guarantee a minimum revenue to keep them active. Therefore, th newly registered driver partners may have priority while waiting for platform matching, although the driver partners may not be the closest ones. The heterogeneity in ATS operation strategies also challenges related studies at both system/aggregated and individual levels. For aggregated analyses (e.g. demand and regional operation), we should pay more attentions on appropriate spatiotemporal scale before any in-depth discussions on ATS behaviors. On the other hand, driver-related analyses also have more concerns over their heterogeneity, not only differences between part-time and full-time but also their potential matching priority defined by platforms. Moreover, partner drivers are not limited to one single TNC and are free to join multiple platforms. They can logon all available platforms while driving and accept one request order, which is called multi-homing behaviors. This also makes system analyses more complicated and challenging, since it introduces an additional strategic behavior for order choices.

2.4 Driver-Passenger Matching and System Analyses

This section mainly involves a critical process during taxi operation that are matching (or bilateral searching) between passengers and empty vehicles. The outputs of driver-passenger matching are key metrics for taxi system performance, for instance, passenger waiting time and driver searching time. Moreover, we also review common system-level modeling structures that incorporate demand, supply, pricing, and matching. In final, we summarize challenges while extending taxi system anal-

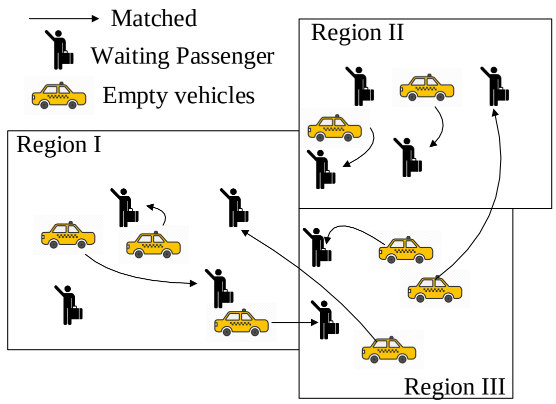


Fig. 2.7.: Illustration of driver-passenger matching in the taxi system

yses to a large scale, generally consisting of numerous individuals and spatially and temporally heterogeneous controls.

2.4.1 Driver-Passenger Matching

As shown in Figure 2.7, the driver-passenger matching process is one empty vehicle driving to passenger's origin for pickup, which is similar as empty vehicle movements. However, the matching process considers dynamics of both passenger generation and empty vehicles in a certain period and region, other than driver-oriented. Modeling the driver-passenger matching process is in general at aggregated level and quantify waiting/searching time with mathematical forms. There are two classical methods, including Cobb-Douglas production function and its variations, and queue theoretic approaches.

The key component of Cobb-Douglas production function is the meeting rate function, considering number of passengers and empty vehicles in one region during a fixed period. The outputs of Cobb-Douglas production function are both average passenger waiting time in equation 2.2 and average empty vehicle searching time in equation 2.3. Yang and Yang made an extensive discussion on the application for taxi bilateral searching problems and recommended parameter calibrations, for instance, $0 < \alpha_1, \alpha_2 \leq 1, \alpha_1 + \alpha_2 > 1$ and positive value for area size related parameter A [58]. However, parameter calibrations are hard to validate empirically and have not reached an agreement, although it has been employed in most taxi system performance analyses [17, 25].

$$w^c = \frac{N^c}{A\left(N^c\right)^{\alpha_1}\left(N^{vt}\right)^{\alpha_2}} \tag{2.2}$$

$$w^{vt} = \frac{N^{vt}}{A\left(N^c\right)^{\alpha_1}\left(N^{vt}\right)^{\alpha_2}} \tag{2.3}$$

where, w indicates average searching/waiting time; N indicates number of passengers or empty vehicles; α_1, α_2, and A are parameters to be calibrated; and superscript c and vt indicate passengers and empty vehicles, respectively.

Qian and Ukkusuri assumed that all empty vehicles and passengers are evenly distributed in each zone and simplified the Cobb-Douglas production function by removing α_1 and α_2. However, the core idea is still same and counts empty vehicles and passengers for searching/waiting time estimations, as shown in equations 2.4 and 2.5 [26]. In addition, the area related parameter A depends on not only area size but also total road length in corresponding region.

$$w^c = \frac{A}{1 + N^{vt}_{out} - N^{vt}_{ex}} \frac{N^c}{N^{vt}_{out}} \tag{2.4}$$

$$w^{vt} = \frac{A}{1 + N^c} \frac{N^{vt}_{ex} + N^c}{N^c} \tag{2.5}$$

where, subscript ex and out indicate excessive empty vehicles in one region and empty vehicles from other regions.

Another type of classical methods is queue theoretic approaches. In taxi system, one region can be abstracted as one queue with single server. The arrival flow can be defined with empty vehicle arrivals to the region and service rate are passenger pickup rate. Taking the simplest $M/M/1$ with First-Come-First-Served for example, we can easily derive the average empty vehicle searching time (equations 2.6 and 2.7) and average passenger waiting time (equations 2.6 and 2.8). More importantly, the queue theoretic approaches can yield the probability density function of empty vehicle searching time, which can not obtain from Cobb-Douglas production function and its variants. However, the method is limited by strong assumptions on arrival flow (e.g. Poisson) and service time (e.g. exponential) distribution. Sometimes, the reality may violate these assumptions, resulting in biased quantifications. The applications of queue theoretic approaches can date back to 1978 by Daganzo to explore the minimum taxi fleet size problem [101]. Then such approaches are extensively discussed by extending to double-ended queues [14,102] and adapting to non-homogeneous Poisson arrivals [103, 104].

$$\rho = \frac{\lambda}{\mu} \tag{2.6}$$

$$w^{vt} = \frac{\rho}{\mu - \lambda} \tag{2.7}$$

$$w^{c} = 1 - \rho \tag{2.8}$$

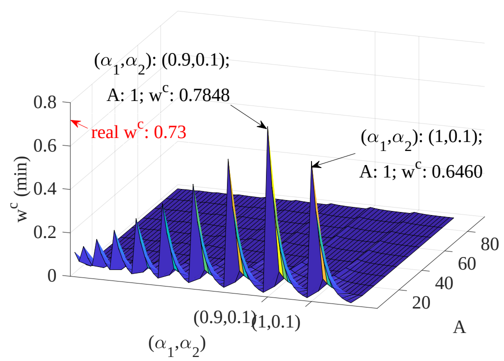
$$W^{vt}(t) = (\mu - \lambda)\, e^{-(\mu-\lambda)t} \tag{2.9}$$

where, λ indicates empty vehicle arrival rate to one region, including both newly online and driving from other regions; μ indicates the passenger pickup rate; ρ indicates utilization rate; and W indicates the probability density function (PDF) of empty vehicle searching time.

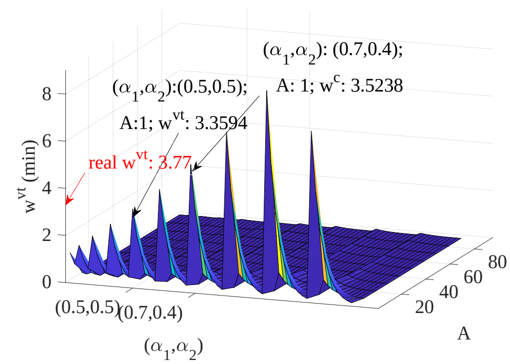
In addition to theoretical comparisons, we also introduce Uber movement records in one community (ID: 226) of Bronx, NYC to empirically examine the method performance. Figure 2.8(a) and 2.8(b) present the estimations under various settings of α_1, α_2, and area related parameter A. With appropriate settings, the estimations on passenger waiting time w^c and vehicle searching time w^{vt} approach to real observations. The appropriate settings are also consistent with Yang's recommendations [58] that are $\alpha_1 + \alpha_2 > 1$. However, the problem is parameter calibrations for both w^c and w^{vt}, since the both estimations require different appropriate settings that are (1,0.1) versus (0.7,0.4). The balance of parameter calibrations for both estimations is still challenging. Moreover, As area related parameter A decreases, in particular for $A < 1$, the estimations are with significant differences even with a very small reduction on A. The setting for A should also be very careful while A is small. Figure

2.8(c) summarize the w^c and w^{vt} changes by different levels of parameter A in Qian and Ukkusuri's method. Different from the original form of Cobb-Douglas production function, the extension has estimations linearly proportional to area related parameter A. However, a similar point is that the both estimations yield one large and one small area related parameter A, respectively. It seems that the passenger waiting and vehicle searching are two different processes and the Cobb-Douglas production function cannot estimate the both very well simultaneously. Figure 2.8(d) and 2.8(e) are changes in estimations depending on vehicle arrival rates λ and passenger pickup rate μ. Although it has relatively larger errors in estimations compared to previous approaches under appropriate settings, the queue theoretic is reliable for both passenger waiting and vehicle searching processes. Moreover, the method can yield more details on distribution of vehicle searching time other than a simple average value. However, the method is also limited by λ and μ. In specifics, it cannot address vehicle oversupply condition under which empty vehicle arrivals are faster than passenger pickups, due to no converges to stationary system state and the estimation results are sensitive to the both λ and μ, in particular the ratio of λ to μ is much less than 1.

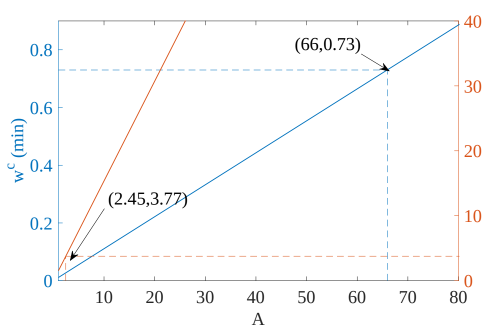
Last, we also have seen two special analyses on empty trip duration by taxicabs or passenger waiting time. Zhang et al. introduced a hazard based duration model with three variations in baseline hazard distributions and explored the impacts of built environment on empty taxi trip duration [27]. Lee and Sohn extended the duration model with additional influencing factors such as demographics, land use, weather, socioeconomics, and transit accessibility [28]. The both analyses only considered factors at origins or/and destinations and ignored factors along the searching route. More importantly, the both studies are individual-driver-oriented and have almost no considerations over demand distribution and passenger waiting time. Meanwhile, Wang and Nie derived a probabilistic model on passengers' waiting time and assumed that the passenger can only be picked up by an idle taxi within an efficient hailing



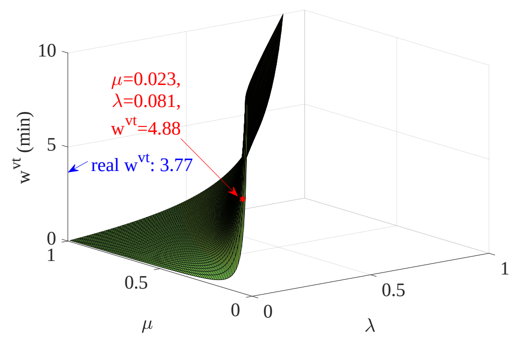
(a) customer waiting time



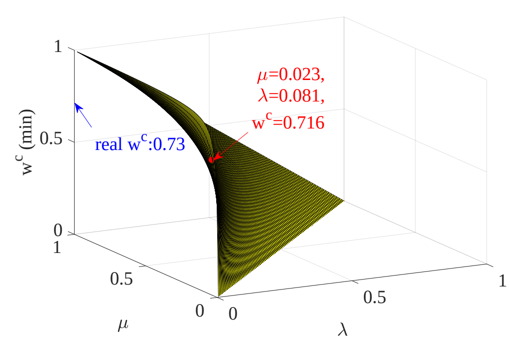
(b) empty vehicle searching time



(c) Cobb-Douglas variants



(d) M/M/1 queue



(e) M/M/1 queue

Fig. 2.8.: Performance of driver-passenger matching methods

distance and that the likelihood of a taxi entering the efficient hailing distance depends on the characteristics of an area [86].

2.4.2 System Performance

Quantifying taxi system performance is critical in taxi system operation, management, and real time control. It combines demand generation and vehicle supply (both movement over road network and online/offline behaviors) with passenger-vehicle matching, and measures passenger waitings, as well as vehicle searching/occupancy. However, modeling such complex interactions is not easy and has three main challenges that are spatial heterogeneity, network externalities, and role of stochasticity.

- Spatial Heterogeneity. Since the taxi activities are highly associated with socio-economic variables, the taxi rides also shows significant spatial heterogeneity, particularly, during rush hours. In other words, most of the taxi pickups may be concentrated in some high economic activity zones, such as central business district, while dropoffs are concentrated in other zones. The modeling has to consider properties that dictate the spatial heterogeneity in an urban area.

- Network externalities. Whenever a customer engages a vehicle, this not only decreases the instantaneous vehicle availability at the source location, but also affects the future vehicle availability at all other locations within short timescales. The impacts of network externalities are more significant under dynamic pricing in ATS. Since changes in vehicle availability can be reflected in dynamic pricing this affects future vehicle supply (e.g. induced supply).

- Role of Stochasticity. In a two-sided market, not only do customers choose when to request a ride but also drivers choose when to work, how long to work, and where to search for customers. Moreover, the platforms frequently examines local network conditions and develop corresponding fare rate for a specific time interval and location. This will in turn affect demand and supply. Even if the transportation network is symmetric (i.e. uniform arrival rates and routing choice at all nodes over a regular grid network), the stochastic nature of arrivals will also quickly drive the system out of balance and hence leads to instability.

The spatial heterogeneity has been discussed a lot, generally addressed with region-specific models on passenger and vehicle arrivals rather than one overall model for one whole city/system. However, the existing studies did not provide any guidelines on the appropriate spatial level for modeling. The remaining two ones have relatively fewer attentions. In specifics, the taxi system interacts with urban road system in a way that more (or less) taxi flows will be more (or less) likely to lead congestion and increase (decrease) taxi travel time. But we have not seen any discussions on the interactions, even simple inclusions of stochastic travel time. The role of stochasticity is also critical in particular during vehicle-passenger bilateral searching. But, most studies are built upon simple assumption (e.g. evenly distributed passengers and vehicles) or classical matching functions (e.g. Cobb-Douglas product function) for average performance metrics (e.g. average waiting/searching time). There lacks stochastic details in the process of vehicle-passenger bilateral searching. Actually, the large-scale and stochastic nature of taxi system makes optimal control critical to understanding the market dynamics. The system modeling should be not only rich enough to capture the salient features of both passengers' and vehicles' behaviors but also the stochastic nature of the demand-supply dynamics and the resulting stability of the system. A quick review of literature is as follows:

The first line of research is the aggregated models that formulate the relationships among system performance metrics by calculating overall demand, supply, waiting time, and searching time and explain major variations in the whole system. Yang et al. derived a system of nonlinear simultaneous equations to measure taxi system demand, utilization, and level of services through introducing number of taxis, fare, income, and occupied travel time as exogenous variables and daily demand, passenger waiting time, vacant taxis, utilization, and average taxi waiting time as endogenous variables [12]. Xu et al. utilized a neural network for complex taxi demand-supply modeling with several exogenous variables (i.e. number of licensed taxis, incremental charge of taxi fare, average taxi ride travel time, average disposable income, and population and customer price index) as inputs, which can lead to estimations of daily taxi passenger demand, passenger waiting time, vacant taxi headway, taxi utilization rate, and average taxi waiting time [15]. Mu and Zhao introduced a queueing theory with both first-come-first serve and random order of service for taxi system modeling and estimated system performance, such as waiting time, cost, occupancy, and profit [13]. Wong et al. presented the taxicab-customer dynamics as a double-ended queueing system and integrated with a taxi movement modeling based on absorbing Markov chain approach that can account for stochastic searching processes of taxicabs [14]. These models are first few studies for taxi system dynamics, but primarily focused on modeling TTS system without any considerations of spatial variations. They employed different modeling structures to explain interactions among the whole system performance. Hence, the major limitation of aggregated models is the capability of addressing stochastic nature of the system and spatial dependencies in both demand and supply.

The second line of research is the game-theory based equilibrium models that investigate dynamics of both drivers and passengers. Based on different definitions of equilibrium, there are three modeling structures: spatial demand-supply equilibrium, competitive equilibrium, and other defined stationary distribution states. The demand-supply equilibrium has many applications in taxi market. Manski and Wright firstly reported an investigation of demand-supply equilibrium in a regulated taxi market, which integrated few aggregated models of demand, supply, taxicab availability, and utilization [105]. Yang et al. presented a stationary equilibrium state that movements of vacant taxis over the network should satisfy the customer demand at all origin zones [71].Yang et al. introduced the Cobb-Douglas type production function for taxi-customer waiting time estimation and derived a stationary equilibrium where demand and number of vacant taxis are same at any specific meeting location [17]. Extending the previous study, Yang and Yang examined the equilibrium properties of an aggregated taxi market, including social optimum, emergence of a Pareto-improving market regime, and market equilibrium under Pareto optimality [58]. Later, Yang et al. extended network equilibrium with congestion externalities through considering normal traffics [106]. Wong et al. adapted the network equilibrium model to the multiple users and modes and derived a simultaneous structure with two equilibrium sub-problems: demand-supply equilibrium of each mode and searching/waiting time distributions [16]. A key problem in aforementioned studies is the prerequisites of perfect market by demand-supply equilibrium, which is impossible in reality. Another group of taxi system equilibrium refers to the concept of Nash equilibrium in the field of game theory. Yang et al. proposed a multi-period taxi system modeling structure using a network presentation and derived an equilibrium that taxi drivers cannot increase individual profits by altering individual's working hours [19]. Wang et al. modeled the taxi system at Nash equilibrium -a well known concept in economics - as a Stackelberg game between maximization of social welfare by city agencies and profit maximization by taxi firms and drivers through proposing a bi-level programming [18]. Wong et al. proposed a bi-level programming where the lower level for taxi and normal traffic assignment procedures, where apply user equilibrium[7], and the up level for taxi movements and thus included the congestion effects into taxi system modeling [72]. Zha et al. included ATS driver supply into a bi-level equilibrium model and explored the performance of surge pricing [107]. Qian and Ukkusuri developed a similar bi-level structure with passengers as leaders to minimize travel cost and both ATS and TTS drivers as followers to maximize their own revenues. Additional studies also investigated the state distribution of taxi system and identified the existence of stationary states. Frechette et al. proposed a competitive equilibrium state determined by a set of distributions, such as demand, waiting time, searching time, and optimal shift starting and stopping [21]. Buchholz separated the taxi system into two subsystems that is one with random customer-driver matching at almost all city streets and another with first-in-first-out service

[7]There are no travelers who can reduce their journey times through changing their routes

at airports and proposed the spatial equilibrium state that was a sequence of taxi system states, transition beliefs, and policy functions over each small homogeneous division [20]. To sum up, the current studies explored much on TTS and have limited explorations on ATS or both. Moreover, the equilibrium model is designed to measure the performance of certain control strategies, for instance pricing and entry regulation, and maximize social welfare, consumer surplus, and producer surplus under a given stationary state, which limits the application for real-time track system performance. Going deep into equilibrium model structures, we can identify multiple weak assumptions of passenger-vehicle matching, independence of taxi flow and travel time over road network, and taxi market with perfect knowledge. These assumptions are neither empirically validated nor without appropriate descriptions on stochastic nature.

The third line of research is significantly different from the previous two lines, since it focuses on individual passengers and drivers other than a group in one defined spatiotemporal scale. The studies generally emulate individual behaviors and interactions with each other by describing system participants or rides as agents over road network or nodes in a graph. The first option is the agent based simulation that models each participant as an agent and learns taxicab-passenger dynamics as rules. Grau and Romeu assigned passengers to nearest vacant taxicabs and simulated taxi system with agent based model for discrete passenger generation, vehicle movement, and passenger pickup events [108]. Maciejewski and Bischoff introduced two heuristics- nearest taxicab-passenger matching and demand-supply balancing - into the current available agent-based simulation tool for taxi service simulation [23]. The second option is the graph theory that generally present participants and taxicab-passenger interactions as nodes and links, respectively. Zhan et al. proposed a bipartite matching on a graph converted from occupied taxi trips to find a perfect matching of minimum matching cost between passengers and customers [22]. If the first two options include many details while defining agents and relations, the computational loads will be very high, requiring additional methods to enhance model scalability. Thus, current cases in the literature are mostly based on the small hypothetical networks. The drawback of related studies is also apparent, which is the matching rules for passenger-vehicle matching. Most studies directly adopt simple distance-based matching rule, since it is hard to describe the numerous matches among individuals as stochastic processes.

The last but also rising line of research is the queueing network approaches for station based transportation system. For example, most studies assume a station-based autonomous taxi/vehicle system where customers arrive at predefined stations for autonomous vehicle rental, drive to destinations, and drop off rental autonomous vehicles. Most of these studies model each station as one $M/M/1$ queue with the assumptions of Poisson arrivals and exponential service then connect queues based on routing [109–111]. The process yields a closed Jackson network with product-form stationary distribution. Queueing networks allow analysts to incorporate two important forms of qualitative prior knowledge: first, the structure of queueing network can be used to capture physical road network connectivity; and second, the queueing theoretic approaches inherently incorporate the assumption that response time explodes when the workload approaches system's maximum capacity, which is useful for examining system performance under heavy flow and worst cases. In addition, multiple studies also introduced both passenger and driver arrival flows and proposed a double-ended or synchronization process queue for passenger-vehicle matching [112, 113]. However, the literature emphasizes approximation techniques to find optimal control policies under queueing network structure, rather than validating queueing network and improving modeling structure for taxi system. On the other hand, queueing network can introduce additional type of queue structure for road network performance, which is useful for quantifying interactions between urban road network and taxi system. There are also alternative models for such interactions. Long et al. applied conservation law into the traffic flows[8] of private cars, occupied taxis and vacant taxis with strategic path-choice behaviors, and measured the taxi performance on utilization and revenues [91]. Ramezani and Nourinejad modeled

[8]A mathematical form of relationships among flow, density, and speed.

the taxi movements in homogeneous spatial units through integrating Macroscopic Fundamental Diagram [9] and derived dispatching strategies with minimal network delays. The results also indicated the advantages of taxi system with dispatching center versus without dispatching [114].

2.4.3 Complex System Optimization

For the last few decades, the taxi market is strictly regulated in terms of fixed fare rates and entry limitations in many cities. Finding an optimal regulation polices, as well as measuring impacts of regulations, are of interest. Yang et al. discussed optimal fleet and fare rate under different regulation scenarios by Stackelberg game [10] and social optimum [71]. These optimums were also considered under an extended environment with congestion externalities by normal traffic [106]. Buchholz estimated the optimal fixed-plus-distance fare structure and measured impacts of free entry, as well as combination of free entry and optimal pricing, using developed spatial equilibrium. Moreover, he discussed the potentials of differentiated pricing [20]. Chang et al. explored the optimal taxi fleet size through a bi-level programming with maximum profit for private sector and minimum cost for public sector in a market with both GPS-equipped and non-GPS-equipped taxicabs [115]. Zhang and Ukkusuri developed two Stackelberg games among taxi agency, drivers, and passengers, and developed a comprehensive travel cost function with external costs such as environmental, safety, and congestion [116]. The core procedure is to propose multiple revenue or cost based objective functions at system level and solve a bi- or multi- level programming for optimal pricing rate, fleet size, or both. This is classical problem in the field of operation research thus has various well known solution approaches.

[9]The Macroscopic Fundamental Diagram is a property of a road network in a homogeneous spatial unit, which relates the network accumulation (or vehicle density) to the outgoing vehicle flows.

[10]The regulator is the leader who has control of either fare or entry. The taxi firms are follower who responses to leader's decisions and decides whether to enter the market or choose a fare to maximize profit.

As the rise of ATS, there emerges more discussions on ride-sharing system targeting a higher vehicle occupancy rate. The new sharing system involves a service provider that matches drivers and passengers with similar itineraries allowing them to travel together and share the costs, which are thought to be an efficient and cheaper taxi-similar service. Agatz et al. summarized state-of-art and recent progresses on dynamic ridesharing, presented all related problems in ridesharing modeling, and showed few potential directions, for instance, fast optimization approaches, incentive schemes, and inclusion of users' strategic behaviors [117]. Furuhata et al. investigated the existing ridesharing system and presented some specific challenges, including design of attractive mechanisms (instantaneous price quote, incentives, and truthfulness), a concierge like ride-matching (users' preferences, multi-hop rides, and consolidation with other modes), and building social trust among unknown travelers [55]. Also discussed in the above review paper, the ridesharing problem is proved to be a NP-hard problem, involving both ride-to-ride matching and vehicle assignment to rides. There are various heuristic approaches for such problem, including a heuristic approach that can address routing planning and vehicle-ride matching [118, 119], a decremental approach that can efficiently search closest taxi based on the idea of gradually reducing the search area [120], and a ride-matching algorithm based on clustered ODs [121]. However, these heuristic approaches are generally built upon simple underlying assumptions. Ozkan and Ward challenged the closest vehicle assignment policy and proposed a new dynamic matching scheme based on continuous linear program that can account for spatial effects in demand and supply [122]. Agatz et al. formulated the ridesharing problem as a sophisticated optimization structure instead of simple greedy ride matching rules and confirmed the outperforms [123].

The rising ridesharing problems are challenging, compared to taxi regulation policy problems. Since the ridesharing explores at individual other than aggregated level, and involves numerous individuals (also numerous objective decision variables and constraints) in a large-scale system. In addition to the computational difficulties in large-scale optimization problem, the variety of ridesharing designs and passenger requirements while matching also increases the degree of difficulty in models and solutions. On one hand, the variants of ridesharing designs are proposed depending when and where to pickup multiple passengers. Schindele proposed a novel ridesharing modeling structure that an occupied taxicabs with remaining seats can pick up passengers en route to its destinations, in addition to passengers matching at origins [124]. Tao and Chen described the taxi ridesharing trial in Taipei and presented the heuristic dynamic ride matching schemes that the closest driver pick up the first passenger then other en-route passengers for the cases of one origin to many destinations and many origins to one destination [125, 126]. Lee et al. also addressed the many-to-one and one-to-many ridesharing problem but with a two-stage approach including ride matchings and taxicab assignment based on minimum pickup costs [127]. Ma et al. devised a large-scale dynamic ridesharing algorithm that firstly match a taxicab with one user then insert other passengers with minimum additional incurred travel distance, which was improved by a lazy shortest path calculation strategy to reduce computation load [128]. Two additional studies also introduced the meeting points for picking up shared passengers instead of original origins. Ma and Wolfson proposed a new slugging form of ridesharing that taxicabs will pick up passengers at a closest location other than their origins and solved a delay bounded and capacitated slugging problem [129]. Stiglic et al. also investigated the impacts of meeting points instead of original origins and destinations in ridesharing system and demonstrated the significant increase in matched rides and decrease in system mileage [130]. The last two studies identified the ridesharing system with round trips and parcel transport. Lloret-Battle et al. addressed the ridesharing system with ride-back that ensures a backing trip fro matched riders based on maximum-flow-minimum-cost optimization problem [131]. Li et al. discussed the potentials of shared taxis with both people and parcels with two mixed integer linear programming [132]. On the other hand, existing studies also addressed various passenger requirements, for instance, equity, mobility, and environmental benefits. Ma et al. quantified both drivers' and passengers' benefits from ridesharing and identified an optimal discount rate resulting in equal benefits [133]. Wolfson and Lin et al. addressed the fairness problem while seeking optimal ride matching solely based on cost savings, arising from the conflicts between self-optimization and system optimality [134]. Chen et al. derived a dynamic ridesharing schemes in vehicular ad-hoc network that can save fuel and reduce pollution [135]. Yu et al. empirically measured the environmental benefits of DiDi ridesharing services in Beijing and indicated substantial energy savings and emission reductions [136]. Xu et al. developed a solution to traffic equilibrium with ridesharing that can evaluate impacts of ridesharing on traffic congestion and vice versa [137]. Yousaf et al. investigated the multi origin-destination path planning problem using a unified framework with multiple conflicting objectives, which can enable the flexibility for drivers to get sub-optimal paths according to their own requirements [138]. The study was further improved through incorporating users' preferences on socially close partners and drivers' preferences on various feasible paths into the state-of-art ride matching algorithms [139]. Ye et al. introduced the non-dominated sorting genetic algorithm to address the multiple objectives in the problem of taxi ridesharing routing optimization [140].

In addition to NP-hard optimization problem and corresponding heuristic approaches, there are two additional solutions for dynamic ridesharing that are agent-based simulation and graph theory. Nourinejad and Roorda proposed an agent based model to address ride matching in dynamic ridesharing problem, which can process multi-passenger matching problem [141]. Martinez et al. utilized agent based simulation to match potential rides that can be combined together, evaluated the potentials of shared taxi system in Lison, Portugal, and indicated the significant fare and time savings to passengers but not jeopardizing that much the taxi revenues [142]. However, the agent based modeling structures are not scalable and are not capable of introducing complicated passenger-vehicle matching mechanisms. The graph theory is further introduced to address the scaling problem. Santi et al. measured the economic and environmental benefits from a ridesharing system by taxi shareability network, where nodes represent taxi rides and links express the potentials of shared ride based on spatiotemporal distance. The graph representations translate a traditional NP-Hard optimization problem into a graph-theoretic framework [143], followed by Altshuler et al. [144]. Alonso-Mora et al. presented a scalable ridesharing algorithm based on a two-stage framework, which can match rides by a pairwise shareability graph and then assign matched trips to vehicles with an integer linear programming [145]. Qian et al. derived an integer linear programming for taxi ridesharing that is also NP-hard and proposed an equivalent graph problem as an alternative solution approach to measure optimal ride matching schemes and incentives for ride sharing services [146].

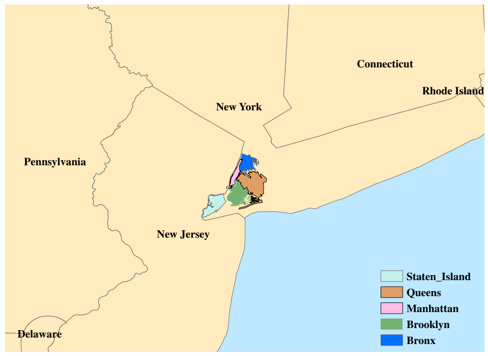
2.5 Summary

In this chapter, we present a review of literature on four major behaviors of disruptive mobility ATS, including demand, supply, surge pricing, and passenger-vehicle matching. The reviews suggest that most of existing taxi system modeling are focusing on TTS-dominated system and fail to capture all strategic behaviors, thus may lead to unreliable results. We find that a detailed representation of strategic behaviors by both ATS and TTS stakeholders and its inclusion into taxi system modeling have potential contributions in understanding urban mobility transformations and developing sustainable control strategies. Although there are emerging studies on all four aspects, the scale of problem is limited to a small neighborhood and modeling structure are challenged by recent progresses in related fields. A review on the existing TTS analyses reveals that the high-resolution mobility dataset enables better understandings of taxi movement patterns, systems, and efficiency enhancements. However, there is almost no such dataset available for ATS.

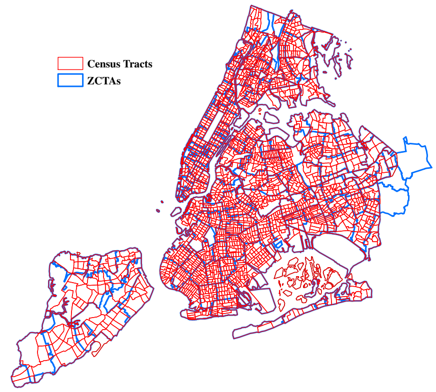
3. LARGE-SCALE MOBILITY DATASET AND AGGREGATION SCALES

3.1 Introduction

Various datasets are required to support explorations in this thesis. First, demand generation problem quantifies how many passengers generate in a given time interval and what impacts from dynamic pricing and other socioeconomic variables. The ATS and TTS ride records are necessary, along with socioeconomic and ATS dynamic pricing at defined spatial and temporal scales. Second, ATS driver stop working model demands a relatively high resolution dataset, including not only the location where one ATS driver drop off passengers but also previous service records by same driver from online till the drop-off. Thus, the detail trajectory data of ATS drivers are recommended. Alternatively, it is also ok to have sequence of empty and occupied trip records by every ATS driver. Third, the dynamic pricing generation model explores interactions between dynamic pricing and demand/supply, thus should provide dynamic pricing multiplier at given aggregated region and time interval, as well as number of passengers and available vehicles. Last, the passenger-vehicle matching and congestion externalities should start from empirical observations on individual matching between one passenger and one vehicle, and vehicle movements over road network, respectively. To sum up, we are demanding very high-resolution mobility dataset for ATS and TTS, as well as ATS dynamic control dataset. More details for each individual and its behaviors are observed, more accurate modeling structures are developed.



(a) Location of NYC



(b) Common spatial divisions of NYC

Fig. 3.1.: Location of data collection area

3.1.1 Study Area and Representative Services

Regarding study area, we select NYC as study area, shown in Figure 3.1(a). This is because: 1) NYC is operating the largest and most representative TTS system with more than 13,000 yellow taxicabs and thousands of green taxicabs in the North America; 2) There are also multiple TNCs launching their services for few years and gaining considerable market shares; 3) NYC has fantastic open datasets that are publicly available, for instance, TTS taxi ride records, land use, and socioeconomics; and 4) NYC has various spatial divisions, from census tracts, zip code tabulation area, community districts, to boroughs, shown in Figure 3.1(b), which makes homogeneous region identifications much easier. Among ATS and TTS are in operation in NYC, we select yellow taxicabs and Uber as representative ones. Since they are most successful in corresponding services and have majority of market share.

3.1.2 Source of Mobility Dataset and Current Progresses

On the side of TTS, NYC taxi and limousine commission has been releasing trip-level ride records from 2009. Each ride is described with origin and destination information, departure and arrival time, travel distance/time, and charged fares. This dataset can reveal perfect knowledge over demand generation. However, it removes several id information, for instance, vehicle id and driver shift id, which prevents us from deep investigations on taxicab supply. To overcome deficiencies in supply information, we also refer to another open dataset with id information, enabling us to track empty taxicab movements and supply level. The open dataset is provided by one third party (ref:https://databank.illinois.edu/datasets/IDB-9610843), who foiled from NYC taxi and limousine commission.

However, on another side of ATS, we can not directly obtain such dataset from government agencies, since almost all TNCs are private commercial companies and not strictly regulated by taxi agencies. Sometimes, TNCs may release ride information for academia but definitely do not share fare and vehicle information. The literature proposed three types of ATS data collection, including survey-based dataset acquired through interview or questionnaires, administrative dataset or ride records released by TNCs, and scraped dataset from platforms. The first two datasets have inherent drawbacks while exploring detailed representations of large-scale system dynamics. For example, the survey-based dataset is difficult to collect under larger spatial and temporal scales, limited by resources and the administrative dataset is generally aggregated at city level thus loses many details of individual behaviors. In addition, the ride records focus on passengers without any information on supply or driver partner side. Thus, our proposed data acquisition method primarily follows the third one that can not only capture both passengers' and drivers' movements but also support efficient data collection in a large-scale system, such as metropolitan areas.

3.1.3 Aggregation Scales

With high-resolution ATS and TTS mobility dataset, another important concern is the selection of appropriate aggregation scales. Compared to one individual discrete event (e.g. passenger generate at one time and location), we are more interesting in a group of events within a defined spatial region and a fixed time interval. The aggregation may be summation or average over all observations falling inside one region and time interval. For example, the demand and supply generation problems emphasize how many passengers and vehicles in total generated from one region every time interval (e.g. 1-hour or 30-min). The dynamic pricing problem shed light on average dynamic pricing multiplier in one region and a fixed time interval, rather than each individual location and every second. Thus, identifying appropriate spatial and temporal aggregation scales is also critical before modeling.

The objective of this chapter is to not only obtain high-resolution ATS mobility dataset, but also propose appropriate spatial and temporal aggregations. Toward this goal, we first develop a framework for large-scale mobility data collection on Uber platform. Then we investigate raw scraped dataset, extract information of interest, and validate with existing official records. Beyond the datasets, we employ multiple statistical hypothesis testing methods to determine aggregation scales, under which the aggregated observations can be assumed with classical distributions. To the best of our knowledge, there are no any prior such work, deeply investigating spatial and temporal aggregations.

3.2 Data Acquisition for ATS

3.2.1 Data collection schemes

Although Uber provides access to some official datasets and APIs to obtain data for research purposes, the data obtained from such tools are inadequate to model the rich spatiotemporal patterns of driver availability, trip characteristics and surge price dynamics at a city-scale market. Specifically, the data collected using official tools has following limitation: 1) The officially released datasets have limited operational statistics at a sparse spatial and temporal coverage, which prohibits in-depth analysis; and 2) the official APIs only provide the trip travel time, waiting time, and estimated fares for a predefined trip but no information on real-time driver availability and surge pricing multiplier (SPM) at different locations. Also, the hourly requests to the Uber API are limited to 2,000, which is inadequate to measure the dynamics at a city-scale area.

Considering the aforementioned limitations, we develop a novel web crawler from scratch to collect the high-resolution spatiotemporal data of available drivers, SPM, and estimated time to arrival (ETA) in real-time in NYC. The web crawler is designed to emulate Uber client app which sends pingClient messages[1] from simulated users' geolocation to Uber server, and receive feedback message every 5 seconds. The feedback message consists of the (near) real-time location of up to eight closest available vehicles, the in-effect SPM and ETA as shown in Figure 3.2.

Note that the tool does not simulate actual ride request but only browse for available vehicles in the vicinity of the data collection location. Moreover, we do not hack any Uber users' app and infringe on their privacy, but only collect operational data, e.g. pricing and vehicle availability. Thus our web crawler tool operates in an ethical manner without disturbing the market supply-demand dynamics.

Since the feedback message from the Uber server consists of location information of up to eight nearest drivers only, we set up our data collection locations densely and send pingClient messages frequently to collect the available driver and SPM data for a large-scale city like NYC. We design the following two tests to find the optimal spacing between the data collections stations and the number of accounts needed to be installed at each data collection location to collect good-quality data using the least number of simulated user accounts.

[1] A pingClient message contains the tentative ride request location, which is not an actual ride request.

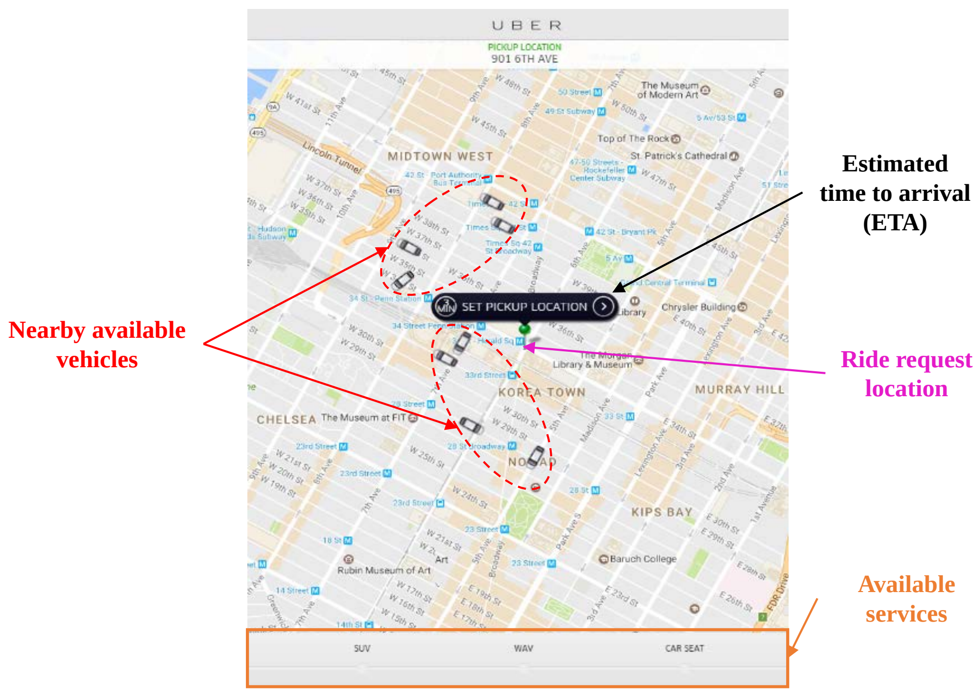
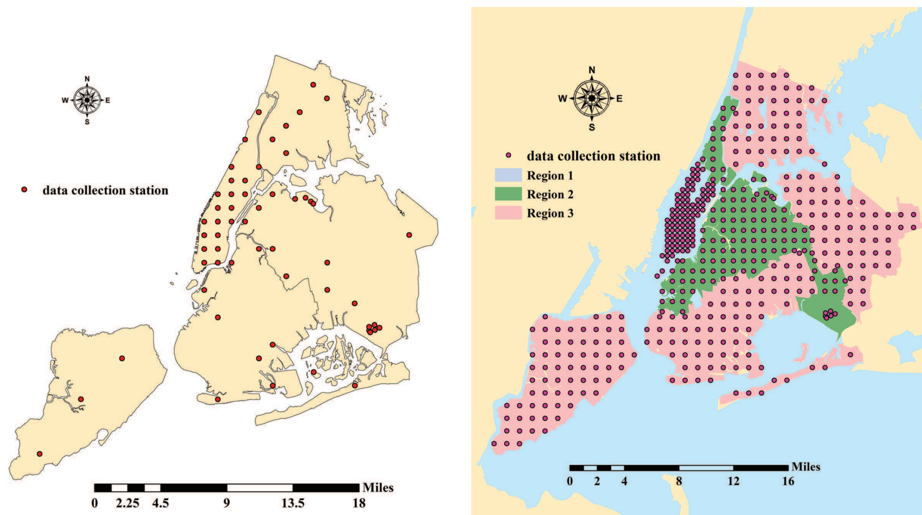


Fig. 3.2.: Uber client app snapshot

Test 1: How many simulated user accounts to set up at each geolocation?

The first question we encounter while developing the web crawler tool is: "whether two pingClient requests sent from a same geolocation can receive same responses (i.e. the same set of nearby vehicles)?" This information would help us set the optimal number of simulated user accounts at each data collection location. For this test, we deploy 58 data collection stations distributed randomly in each of five boroughs in NYC, as shown in Figure 3.3(a). Note that Manhattan has more data collection



(a) 58 data collection stations for Test 1 (b) 470 data collection stations in NYC

Fig. 3.3.: Data collection stations deployment on Uber platform

stations due to higher driver availability. On the contrary, remote boroughs (e.g. Staten Island) have fewer data collection stations. We assign two simulated users at each of the 58 data collection stations, which send pingClient messages simultaneously to the Uber server every 5 minutes for a day. In more than 99.99% of the cases, the feedback messages from the two unique accounts at the same location are exactly same. From this observation, we conclude that only one simulated user account is required at each data collection station to track all nearby available vehicles.

Test 2: What is the optimal spacing between the two data collection stations?

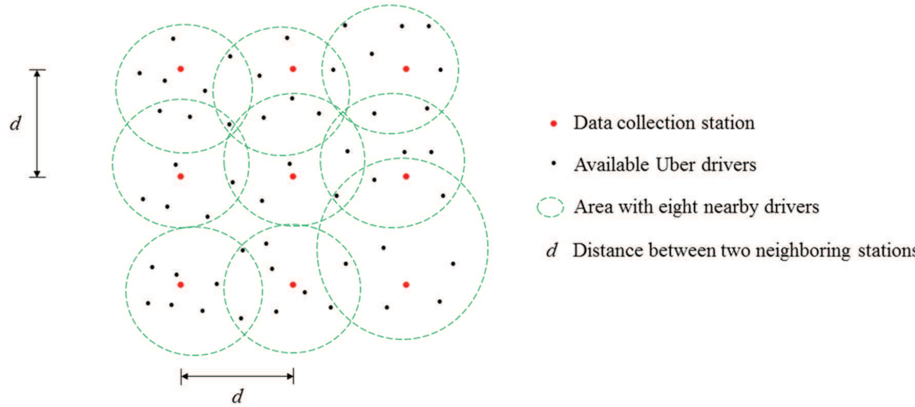
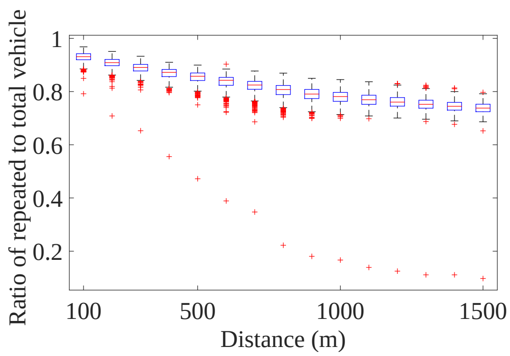


Fig. 3.4.: Illustration of one test-bed with distance d between neighboring stations

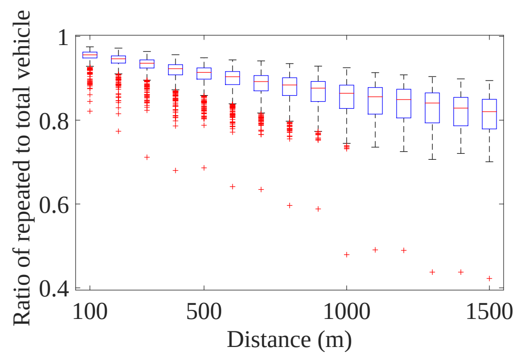
This test is designed to find the optimal number of data collection stations and their locations across the NYC area to (hopefully) capture the entire Uber fleet. Deploying the data collection stations at the vertices of a rectangular grid, we aim to find the optimal spacing between the neighboring stations to have minimum overlap and no gaps in the coverage region. Too large a spacing will fail to cover the entire area of interest and too small a spacing will require a large number of simulated Uber users, which risk being caught by the Uber server.

In this test, we divide the NYC into three regions according to historical taxi rides. These three regions are: region 1 - downtown and midtown of Manhattan; region 2 - upper Manhattan, downtown of Queens and Brooklyn, and the two airports (La Guardia and JFK); and region 3 - other remote areas, as shown in Figure 3.3(b).

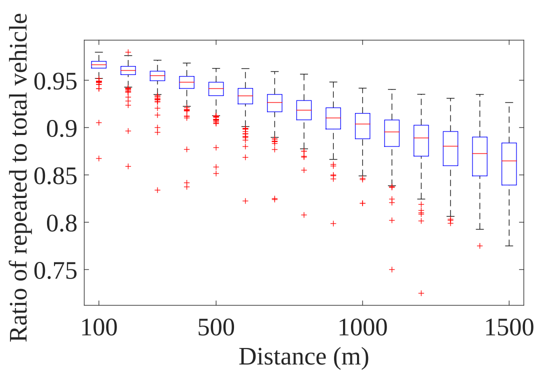
In each region, we setup a test-bed consisting of a 3×3 grid of data collection stations with the distance between neighboring stations being d, as shown in Figure 3.4. Each data collection station (red dot) and its 8 nearest vehicles (black dots) are shown enclosed by a green dashed circle. We use the ratio of number of repeated vehicles (black dots in the overlapping area of two green dashed circles) to the total number of observed vehicles in the 3×3 grid of data collection stations to determine the optimal value of d for each of the three regions. The smaller the ratio is, less is the possibility of capturing same vehicle by two neighboring data collection stations, however the chances of missing some vehicles is larger, and vice versa. We experimented with different values of d, varying from 100m to 1,500m with an increment of 100m and track nearby vehicles every 5s in a day. Figure 3.5 shows the box-plot of the



(a) Region 1



(b) Region 2



(c) Region 3

Fig. 3.5.: Box-plot of the ratio of number of repeated vehicles to the total number of observed vehicles for different values of d

ratio of the number of repeated vehicles to the total number of observed vehicles for UberX in the three regions. We use the optimal value of d, which provides the worst case ratio of the number of repeated vehicles to the total number of observed vehicles to be 0.4 for optimal balance between data quality and the number of data collection stations. Based on this threshold, we use $d = 600m$ for region 1, $d = 1,200m$ for region 2, and $d = 1,500m$ for region 3.

Table 3.1.: Summary of Uber products and the Yellow taxicabs in NYC

Product (product id)	Capacity	Fare			Description
		Base ($)	$/miles	$/minute	
Black (4)	4	7	3.75	0.65	Luxury sedan
SUV (14)	6	14	4.5	0.8	Luxury SUV
UberX (39)	4	2.55	1.75	0.35	Low-cost
Rush (694)	NA	3	4	0	Delivery
UberXL (912)	6	3.85	2.85	0.5	Low-cost SUV
Family (917)	4	2.55	1.75	0.35	With baby seat
POOL (2083)	2	2.55	1.75	0.35	Share and split
WAV (10000936)	4	2.55	1.75	0.35	Weelchair friendly
Yellow taxicab	4	2.5	2.50	0.50	TTS

3.2.2 Dataset Descriptions

Finally, we deploy 470 data collection stations across the NYC, including 99 in region 1, 119 in region 2, and 252 in region 3 as shown in Figure 3.3(b). We deploy 3 and 6 data collection stations at the La Guardia and JFK airports receptively, without considering the $d = 1,200m$ for region 2, due to high vehicle availability at these locations. At each of the 470 stations, we collect the data of available vehicles for the eight Uber products offered in the NYC area. These products (along with the NYC Yellow taxicabs) are summarized in Table 3.1.

We collected the data from April 7, 2017 to May 1, 2017, from 5 AM to midnight each day (a total of 25 days). Each of the 470 data collection location send a pingClient request every 5 seconds and record the response from the Uber server. We do not collect data from midnight to 5 AM since 1) we require few hours to simulate Uber users and generate login authorizations; and 2) there are relatively fewer rides during these hours.

Using the raw feedback messages from the Uber server, we extracted two datasets. The first is the available vehicle trajectory dataset consisting of 1.5 billion time-stamped geolocations and driving direction (course) of available vehicles for different Uber product (dataset size $\sim$ 100GB). Table 3.2 gives a few example of the vehicle

Table 3.2.: Available Uber vehicle trajectory dataset

vehicle id	product id	epoch	course	latitude	longitude
...	...	...	...	...	...
9582b33d4306cd893873cb9ce850fc8047fe6292	39	1491642330881	198	40.67993	-74.00515
9582b33d4306cd893873cb9ce850fc8047fe6292	39	1491642335043	311	40.6801	-74.00573
...	...	...	...	...	...
4346c5a6a508479d2a872665072726a7e6bd2994	14	1491671960662	NULL	40.65519	-73.79486
...	...	...	...	...	...
44af658b8c4ee3c9633f5226a986c11c96636693	39	1491663796544	280	40.77115	-73.86631
...	...	...	...	...	...

trajectory dataset samples, where epoch is the number of milliseconds that have elapsed since January 1, 1970 (midnight UTC/GMT), not counting the leap seconds. The second dataset contains information about the SPM and ETA for different Uber products at different data collection stations every 5 seconds. Table 3.3 shows a few

Table 3.3.: SPM and ETA dataset

station id	product id	epoch	SPM	ETA (sec)
...	...	...	...	...
3	39	1491655029436	1.4	448
3	912	1491655029436	1.4	236
...	...	...	...	...
234	917	1491690726371	1.0	105
...	...	...	...	...

records of the SPM and ETA dataset.

3.3 ATS Data Preprocessing

The Uber trajectory dataset includes the records of online drivers. Note that a driver is online if he/she is available for serving riders, not those who are currently serving a ride. The supply of drivers at a particular location at a given time can be easily estimated using the high-resolution online vehicle trajectory data collected using our web crawler. Unfortunately, there is no information about if the driver is serving a trip or whether he/she went offline. However, we can leverage the short breaks in the online vehicle trajectory data to infer the demand and offline. This is the first preprocessing for ride estimation. On the other hand, we shed light on the active/genuine vs. phantom vehicles. The issue of phantom drivers on Uber platform was first reported by Alex Rosenblat and Luke Stark, a pair of researchers from the *Data & Society* think-tank in an article for Vice's Motherboard [147]. They cited driver and passenger testimonies which suggested that the Uber app shows available vehicles in the passengers' vicinity even when none are there. They also cited an Uber customer-support representative who apparently told a passenger that the app "is simply showing that there are partners on the road at the time. This is not a representation of the exact numbers of drivers or their location. This is more of a visual effect letting people know that the partners are searching for fares." However, the official statement from Uber denies the existence of phantom vehicles and asserts that the cars are real, if occasionally lagging behind real time [148]. In the second data preprocessing, we aim to verify the "phantom driver" theory using data-driven analysis and filter out samples of genuine vehicles.

3.3.1 Ride estimation

For this, we sort the vehicle trajectory data for each vehicle id in increasing order of time epoch. For two consecutive data points, let Δt be the time difference (in seconds) between the two epochs and let Δd be the straight line distance (in meters) between the two geo-coordinate locations. We define cruise, trip, and offline status according to following rules:

- 1-Cruise : $\Delta t \leq 60$ OR ($60 < \Delta t \leq 5400$ AND $\Delta d < 200$)
- 2-Trip : $60 < \Delta t \leq 5400$ AND $\Delta d \geq 200$
- 3-Offline: $\Delta t \geq 5400$

A driver is considered as cruising if two consecutive geolocations are either less than 1 minute or less than 200m apart. A driver is considered on a trip if the time difference between two consecutive geolocations is between 1 minute and 90 minutes and the straight line distance between them is greater than 200m. If a driver does not send his/her location and time updates for up to 90 minutes, we consider him/her to be offline.

Figure 3.6 illustrates the ride estimation process with simple examples. However, the estimation can be improved at two points: the threshold settings and recognition of sequence of 2-3 (i.e. trip followed by offline). The current threshold settings are arbitrary values, derived from our observations on traditional taxi characteristics and other relation studies. The settings may filter out few distant trips and include few cruise intervals as trips. On the other hand, if one driver exit system immediately drop off passengers, we can not distinguish the trip from offline. Since our dataset only contains trajectory point of available vehicles and we do not know when and where the driver drop off passenger.

3.3.2 Samples of Genuine Drivers

For this analysis, we partition the driver ids on the basis of the number of days they were online. For the 25 days of collected data, the number of days a driver id is online varied from 1 to 25. Figure 3.7(a) shows the histogram of driver ids by the number of days they were online. The number of driver ids which were online for just one day is quite high (approximately 3 order of magnitudes higher than the number

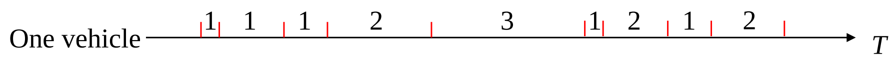
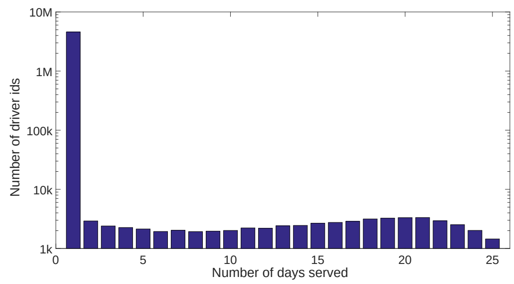


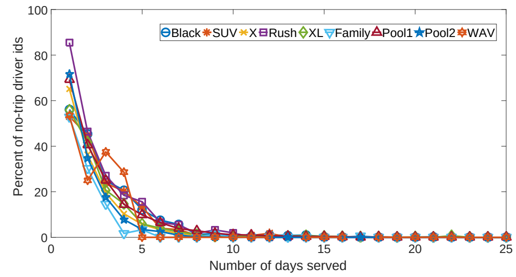
Fig. 3.6.: Illustration of Uber ride estimation based on vehicle trajectory

of driver ids which were online for more than one day). Two possible reasons for this observation could be: dynamic driver id allocation or presence of a large number of *phantom drivers* on Uber platform. Existence of a large number of driver ids who were online for numerous days cast a doubt over the dynamic driver id allocation scheme. This left us with the phantom driver theory.

We found that a large fraction of the driver ids who were online for a fewer number of days (< 6) did not service any trip. They went from cruising to offline without serving any trip (we call them "no-trip drivers"). Figure 3.7(b) shows the percentage of no-trip drivers among all the drivers who serviced that product for those many days for all the Uber products. The percent of no-trip drivers is quite high for drivers who serviced for less number of days (< 6), e.g., the no-trip driver percent is as high as 85.5% for UberRush drivers who served for one day only. As the number of days a driver is online increases, the number of no-trip drivers reduces sharply and there is hardly any no-trip driver among those drivers who serviced for more than 10 days. This behavior of Uber drivers to went from cruising to offline without servicing any trip seems suspicious since the aim of drivers is to earn money by servicing the passengers. If only some drivers exhibit this behavior it can be understood as they may have some emergency and have to log off, but for such a large number of drivers to do so seems like a systematic issue is at play here. As the number of days a driver id is online for increases, almost all of the available drivers service at least one trip indicating genuine driver behavior. This further reinforces our belief that there are many phantom drivers shown to the customers by the Uber app. One point to note about the data processing framework for ride estimation is that we are not sure if a driver can go offline directly after serving a trip or would he/she be online for a few



(a) Number of drivers versus days online



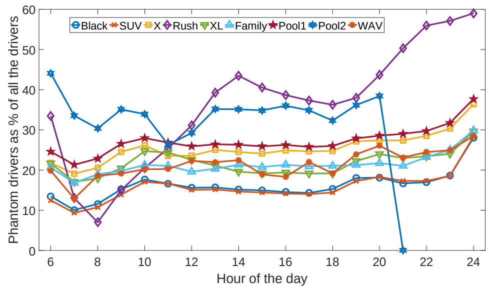
(b) Percentage of no-trip drivers by days serviced

Fig. 3.7.: Pattern of driver online days and the percentage of no-trip drivers

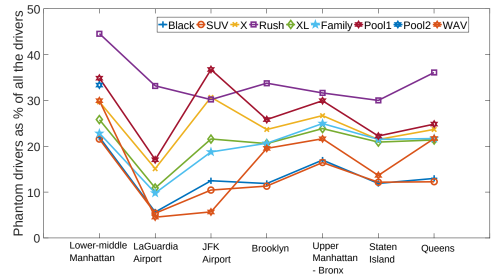
seconds before he/she logs out of the system after completing the trip. If former is the case, one could argue that there are many drivers who would like to serve as Uber drivers for a particular trip they are making themselves such as going to or coming back from work and they went offline after completing that trip. However, since the driver does not know the destination of a potential passenger until they accept the ride request they may not choose the passenger based on their destination (They can cancel the request later but their cancellation rate should not be higher than a small number otherwise Uber might blacklist them). Based on the above discussion, it is fair to assume that if a person is serving as Uber driver, he/she would serve passengers for some time duration (based on their schedule) and would like to take as many rides as possible to make more money. Hence, we conclude that the no-trip drivers are most probably phantom drivers.

Our analysis also found that the number of phantom drivers is a function of the number of genuine drivers in the vicinity, which varies with location and time. Figure 3.8 plots the percentage of phantom driver among all the drivers available for a particular product by the hour of the day and by location (borough).

For most products, the number of phantom drivers lies in the range of 10-30% of all the drivers and increases significantly during evening peak and night hours (8 pm to midnight). This helps provide the riders with a false sense of ample driver



(a) Percentage of phantom drivers by hours



(b) Percentage of phantom drivers by location

Fig. 3.8.: The patterns of phantom vehicles across different time and location

availability even if there are fewer drivers available. For Rush service, this strategy is even more severe since the number of phantom drivers exceeds the actual number of drivers (more than 50% of all the drivers). In addition, since UberX is highly popular in lower-middle Manhattan and there are many (genuine) drivers available, the Uber app usually shows a large number of phantom drivers which sometimes exceeds the number of genuine drivers since drivers from nearby areas can quickly drive to the pick up location and act as a cover for phantom drivers. Comparing by boroughs, the number of phantom drivers is highest in Manhattan as compared to other boroughs due to higher demand and genuine driver availability.

3.3.3 Data Validation

We finally filter out a sample of active drivers who are online for more than 10 unique days and estimate rides. The estimated daily ridership is shown in Figure 3.9. We also compare the estimations with released dispatched rides (about 270,000 daily rides) by Uber, shared by NYC Taxi and Limousine Commission at http://www.nyc.gov/html/tlc/html/technology/aggregated_data.shtml. We observe about 40% to 50% of reported rides. There are few reasons causing the difference: 1) the threshold settings for ride estimation that may exclude few distance trips; 2) the inefficiency in identifying trip-offline sequence that is popular in ATS; and 3) the

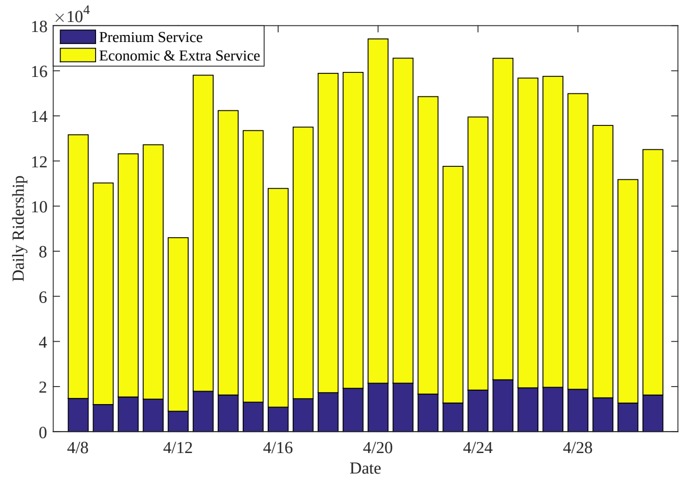


Fig. 3.9.: Estimated daily ridership after data preprocessing

threshold for sampling active drivers that may exclude few genuine vehicles with less than 10-day service.

3.4 Appropriate Aggregation Scales

3.4.1 Statistical Hypothesis Testing

Each passenger or vehicle arrival can be seen as one independent discrete event. Aggregating arrivals at appropriate spatial and temporal scale will result in a sequence of arrival counts at one region of interest during study period. One of simple but classical distribution is Poisson, describing such independent discrete events. Thus, we would also refer to Poisson distribution for taxi passenger or vehicle arrivals. The appropriate aggregation scale is the combination of spatial division and time interval, leading to a condition that taxi arrival events in most spatial divisions can assume to be Poisson distribution.

We employ multiple statistical hypothesis testing method to determine appropriate aggregation scale and statistically test whether the arrival events can be described by Poisson distribution with confidence level of 95%. The null hypothesis is that the observed taxi arrival events under given spatial division and time interval follow Poisson distribution. The alternative hypothesis is that the observed taxi arrival events under given spatial division and time interval do not follow Poisson distribution. The first hypothesis testing method is adapted from Kolmogorov-Smirnov test, considering the application for discrete events [149]. The correction primarily consider sample size, as equation3.1. Once D_n is greater than one critical value from Kolmogorov-Smirnov distribution, it rejects the null hypothesis and the observed arrival events can not be assumes as Poisson distribution.

$$D_n = max_{x \in J} \sqrt{n} |H(x) - F_n(x)| - \frac{1}{\sqrt{n}} \tag{3.1}$$

where, n is sample size; J is levels of arrival count ranging from 0 to maximum count; $H(x)$ is hypothesized cumulative density function of one Poisson distribution with empirical mean estimated from random half of samples; and $F_n(x)$ is empirical cumulative density function for remaining half of samples.

Additionally, we introduce three statistical hypothesis testing method for homogeneous Poisson. The null hypothesis is that taxi arrival events are from Poisson distribution with same rate across time in one given spatial division. The alternative hypothesis is that taxi arrival events are from Poisson distribution but with different rates across time in one given spatial division. Such statistical testing are applied directly on a sequence of arrival counts $\{c_1, c_2, \cdots, c_i, \cdots, c_n\}$ and are generally based on χ^2 distribution. The differences in three methods are in statistics computation. The Anscombe method first transforms original arrival counts then compute statistics with squared errors, as shown in equation 3.2. The likelihood ratio statistics are computed from ratio to mean value, as shown in equation 3.3. And the conditional χ^2 statistics are measured with ratio of squared error to mean value, as shown in equation 3.4. If computed statistics are greater than critical χ^2_{n-1}, it rejects null hypothesis at confidence level of 95% and the taxi arrival events are from Poisson distribution with time-dependent rates.

$$\begin{aligned} y_i &= \sqrt{c_i + \tfrac{3}{8}} \\ T_{anscombe} &= 4 \sum (y_i - \bar{y})^2 \end{aligned} \tag{3.2}$$

$$T_{likelihood} = 2 \sum_{i=1}^{n} c_i ln \left(\frac{c_i}{\bar{c}} \right) \tag{3.3}$$

$$T_{anscombe} = \sum_{i=1}^{n} \frac{(c_i - \bar{c})^2}{\bar{c}} \tag{3.4}$$

where, n is sample size; c_i is arrival count in a specific time interval i and spatial division; and $\bar{c}$ is mean values of all arrival counts.

3.4.2 Empirical Results

The potential spatial scales are mainly based on four administrative divisions in NYC, including Borough (5 in total, ∼60.4 mi^2 on average per Borough), Community Districts (71 in total, ∼4.3 mi^2 on average per community district), Zip Code Tabulation Area [ZCTA] (214 in total, ∼1.4 mi^2 on average per zip code tabulation area), and Census Tracts (2165 in total, ∼0.14 mi^2 on average per census tract). Note that we use administrative divisions rather than grid based spatial scale, since most socioeconomic variables are only available at administrative divisions and modeling will be much easier to include.

We test seven different count intervals, that are 1-min, 5-min, 10-min, 15-min, 20-min, 30-min, and 60-min. In other words, we count TTS, ATS, or both TTS and TTS pickups (or vehicle arrivals) every count interval then test whether the flow can be described with Poisson distribution. In addition, we also test homogeneous period selection, considering time-of-day and day-of-week effects. Regarding the peak (or off peak) hour, we include three difference cases, including 1-hour period (peak: 6pm to 7pm, or off peak: 10am to 11am), 2-hour period (peak: 5pm to 7pm, or off peak: 9am to 11am), and 3-hour period (peak: 5pm to 8pm, or off peak: 9am to 12pm).

Moreover, we classify the weekdays from Mondays to Thursdays, compared to the all seven days case.

Overall, regarding the spatial and temporal aggregation, we prefer community districts and 1-minute count interval for both ATS (i.e. Uber) and TTS (i.e. yellow taxi) in NYC. Empirical evidences are summarized as follows. Note that we only present the hypothesis test results for passenger pickups and vehicle arrivals during off peak hours, since the identified patterns from hypothesis testing are similar between off peak and peak hours.

Figure 3.10 to 3.17 show the percentage of zones not rejecting Poisson distribution

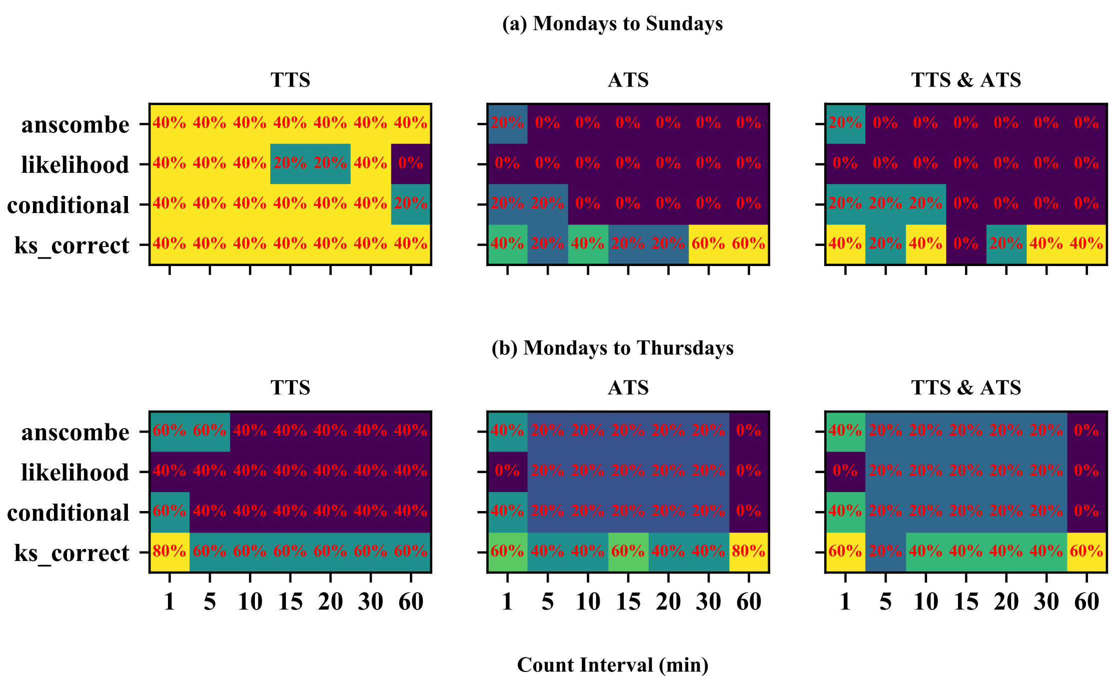


Fig. 3.10.: Hypothesis tests for passenger pickups at Boroughs in one-hour off peak

with four proposed hypothesis testing methods, 7 count intervals, and day of the week. It is apparent that the smaller count interval generally leads to more zones not rejecting Poisson assumptions, across almost all plots. Although several hypothesis tests by corrected KS reveals similar percentages from 1-min to 60-min count interval, the homogeneous Poisson tests generally reject the null hypothesis of arrival counts are

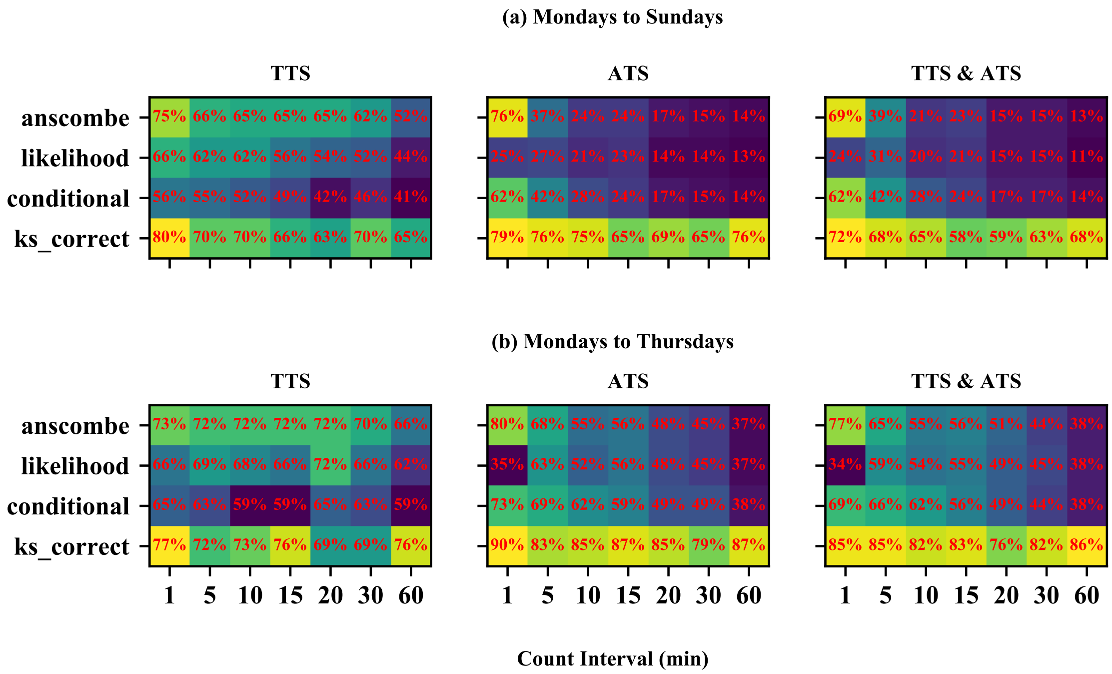


Fig. 3.11.: Hypothesis tests for passenger pickups at Community Districts in one-hour off peak

from one single homogeneous Poisson distribution if we have a larger count interval. Thus, we can select 1 minute interval for both passenger and vehicle arrivals regardless of ATS and TTS.

Figure 3.10 to 3.13 summarize the percentages of zones where passenger pickups can be assumed as Poisson distribution at four levels of spatial aggregations, respectively. Within our expectations, the too large or small aggregation does not have perfect results, since large zones have more spatial interactions and heterogeneity and small zones usually do not have any rides. Both community district and ZCTA aggregations have higher percentages of zones not rejecting Poisson assumption. And community district aggregation generally has a slightly higher percentage, regardless of methods, taxi service types, count interval and day of the week. One interesting point is that the passenger pickups of TTS and ATS are likely independent considering significant Poisson tests for TTS, ATS, and overall pickups of both services.

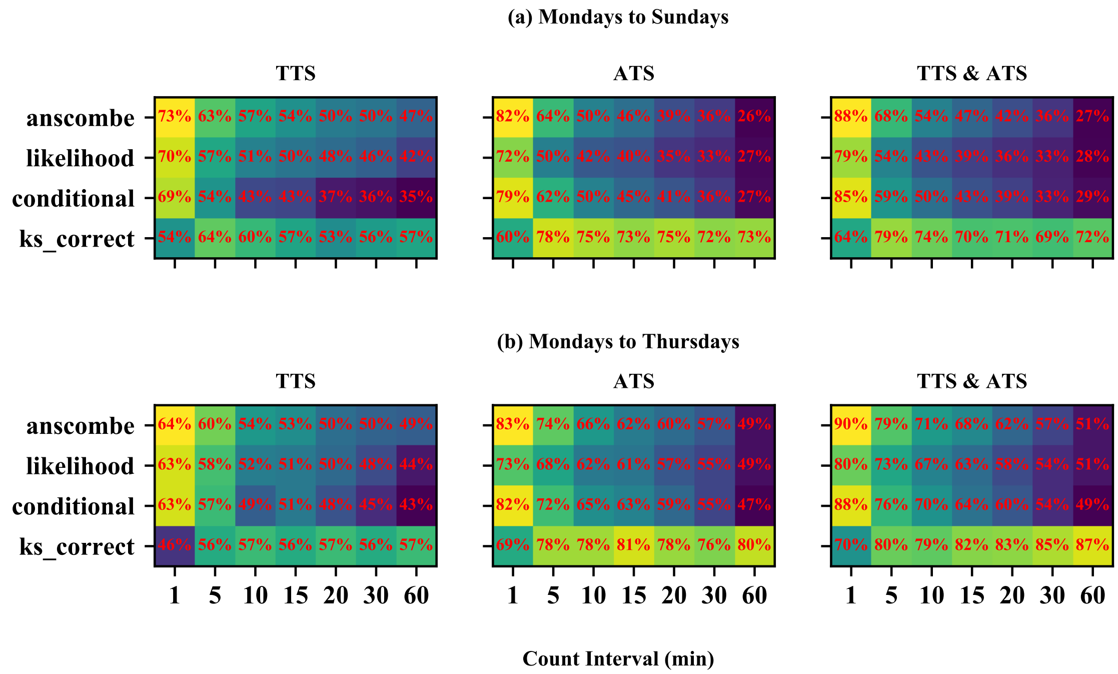


Fig. 3.12.: Hypothesis tests for passenger pickups at ZCTA in one-hour off peak

Figure 3.14 to 3.17 exhibit the percentages of zones where vehicle arrivals (newly online by ATS driver partners or a new shift of TTS driver) can be assumed as Poisson distribution at four levels of spatial scale, respectively. Similar as test results for passenger pickups, both community district and ZCTA aggregation reveals higher percentages. In particular, the percentages are sometimes close to 100%. To sum up, we choose community district aggregation.

Except for the hypothesis testing for Poisson assumption, we also determine the homogeneous study period for passenger and vehicle arrivals, primarily considering time-of-day and day-of-week. Based on our empirical evidences, it is better to limited our study period to one-hour peak (or off peak) hour and weekdays between Monday and Thursday. However, extending study period to more hours and all seven days will only slightly reduce percentages of significant zones. If there are limited observations, extending study period to include more data are also feasible.

Figure 3.11, 3.18 , and 3.19 compare the hypothesis results for community district

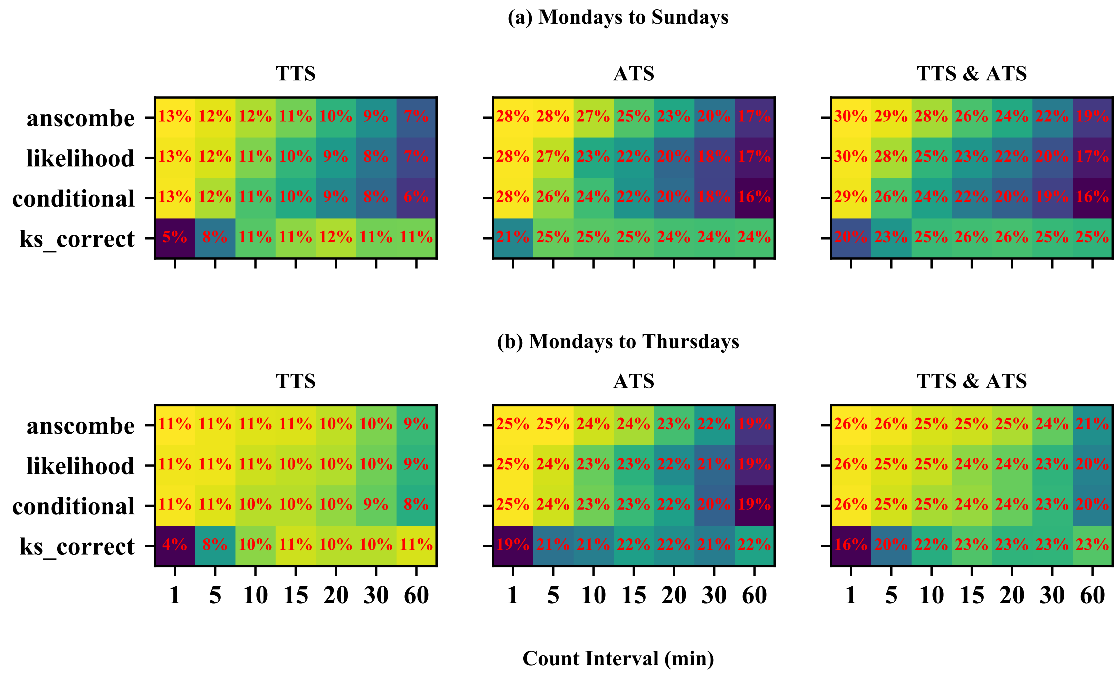


Fig. 3.13.: Hypothesis tests for passenger pickups at Census Tracts in one-hour off peak

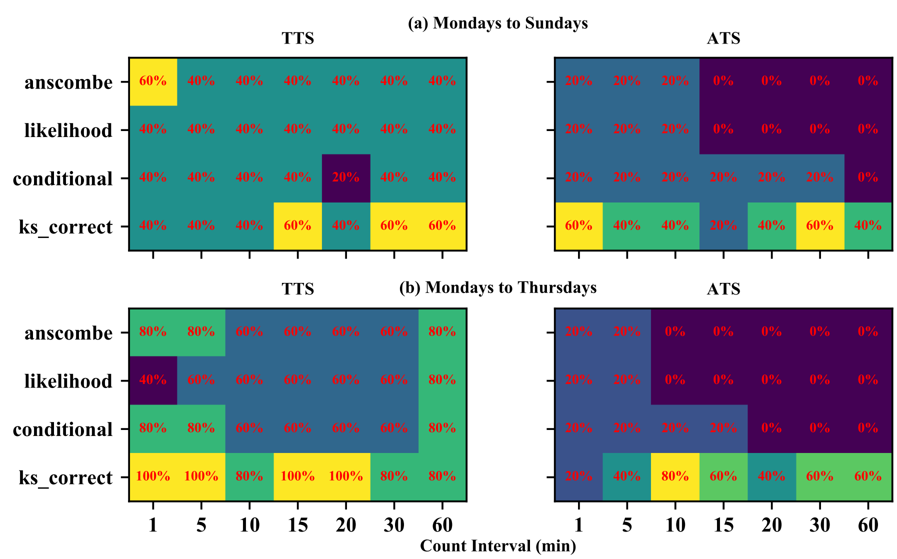


Fig. 3.14.: Hypothesis tests for vehicle arrivals at Boroughs in one-hour off peak

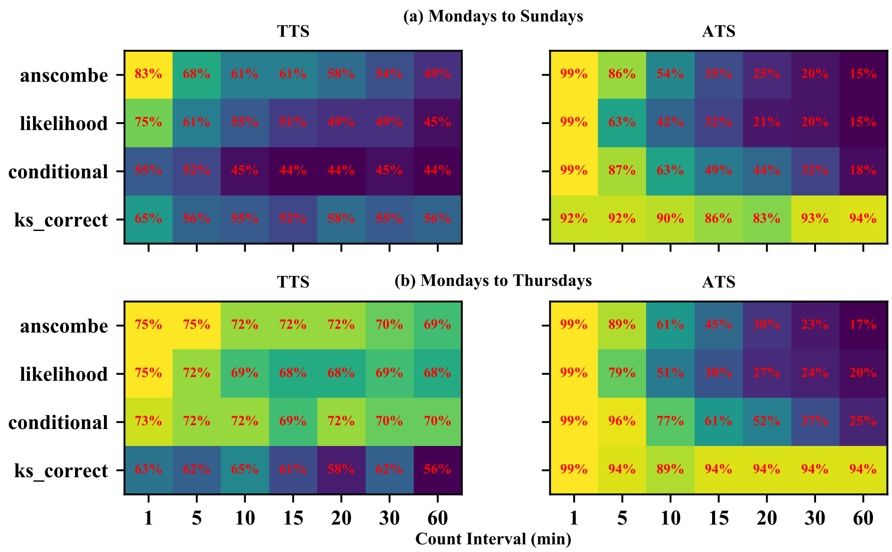


Fig. 3.15.: Hypothesis tests for vehicle arrivals at Community Districts in one-hour off peak

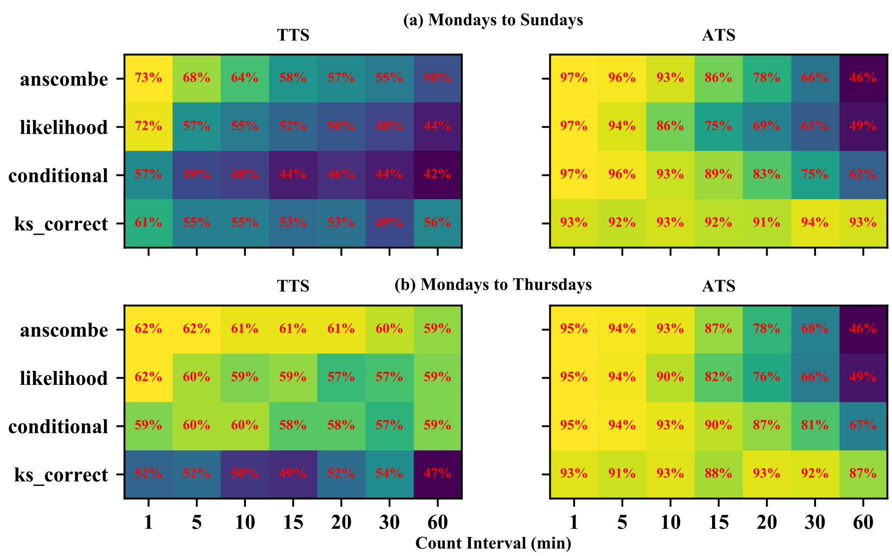


Fig. 3.16.: Hypothesis tests for vehicle arrivals at ZCTA in one-hour off peak

aggregation of passenger pickups across different levels of off peak hours, as well as day of the week. As number of hours included into off peak increase, less community districts are not rejecting Poisson distribution. In addition, limiting to the week-

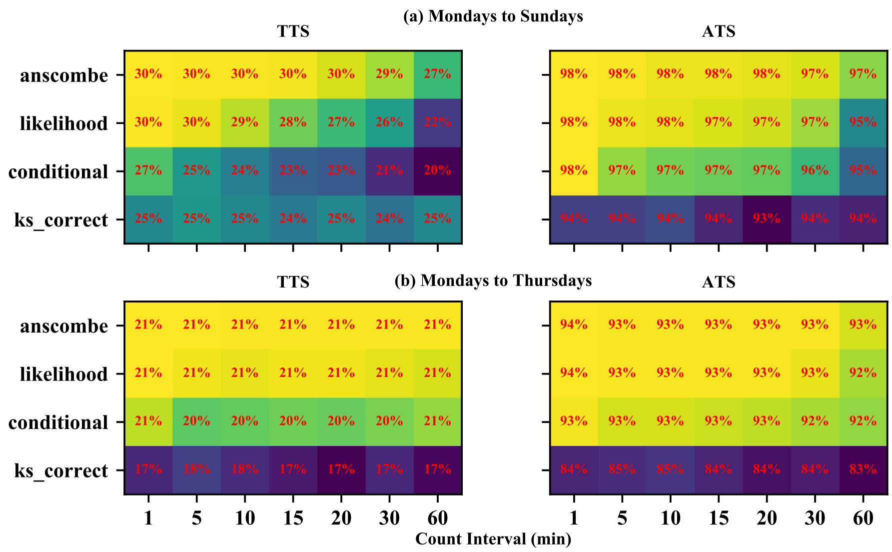


Fig. 3.17.: Hypothesis tests for vehicle arrivals at Census Tracts in one-hour off peak

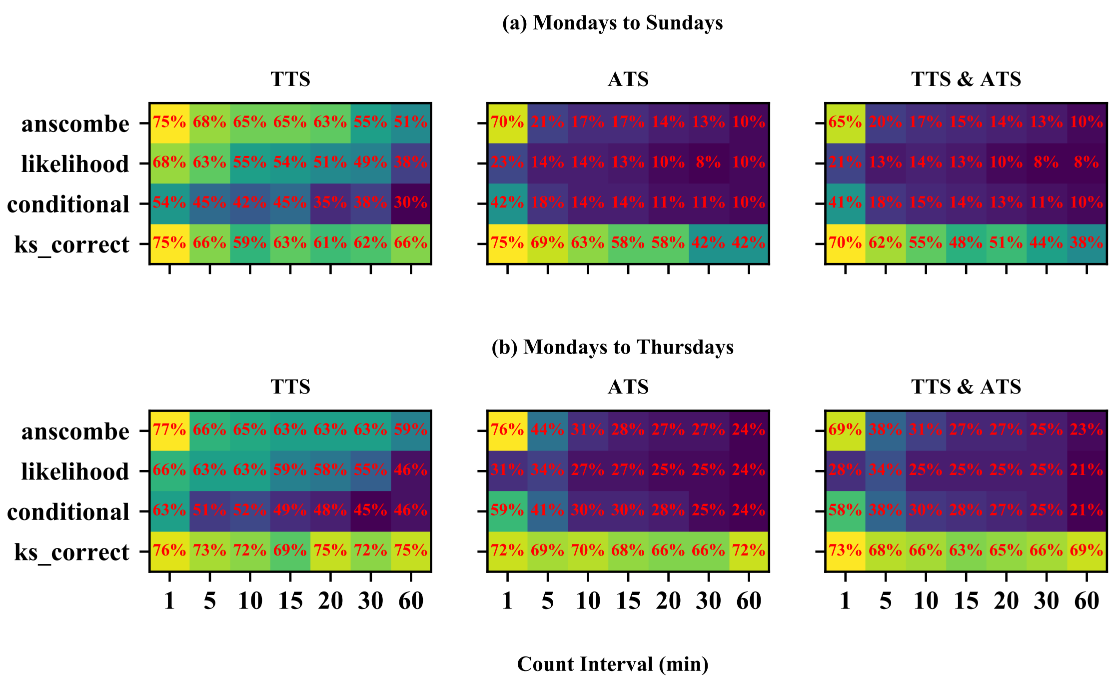


Fig. 3.18.: Hypothesis test for passenger pickups at Community Districts in 2-hour off peak

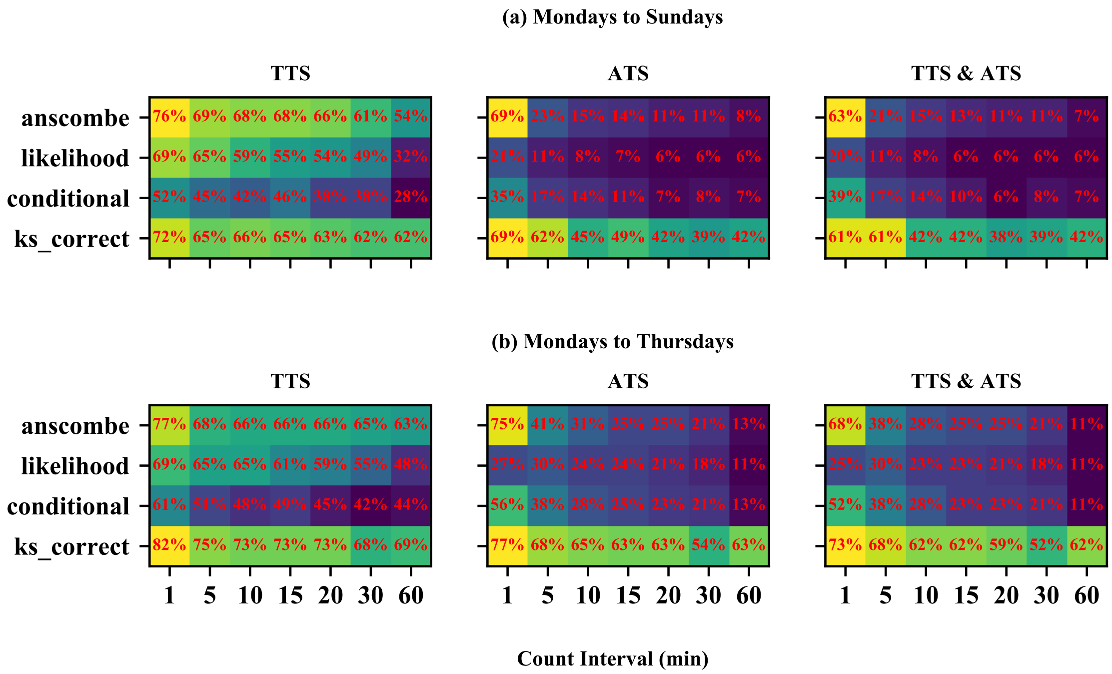


Fig. 3.19.: Hypothesis test for passenger pickups at Community Districts in 3-hour off peak

days from Monday to Thursday can slightly increase percentages of significant zones. Figure 3.15, 3.20, and 3.21 compare the hypothesis results for community district aggregation of vehicle arrivals across different levels of off peak hours, as well as day of the week. There are no big differences in the percentages by week of the day, as well as number of hours in off peak period. However, introducing more hours or focusing on weekdays can lead to very small increases in percentages.

With aforementioned aggregation scales and study period, we find that both passenger and vehicle arrivals can be described with one homogeneous Poisson distribution, in more than 80% of community districts (except for TTS vehicle arrivals with around 65%), as shown in Figure 3.11 and 3.15. In addition, we plot the community districts not rejecting Poisson distribution in Figure 3.22 and 3.23. The both figures indicate that those insignificant (i.e. rejecting Poisson assumption) community districts mainly locate in remote suburban areas, generally with

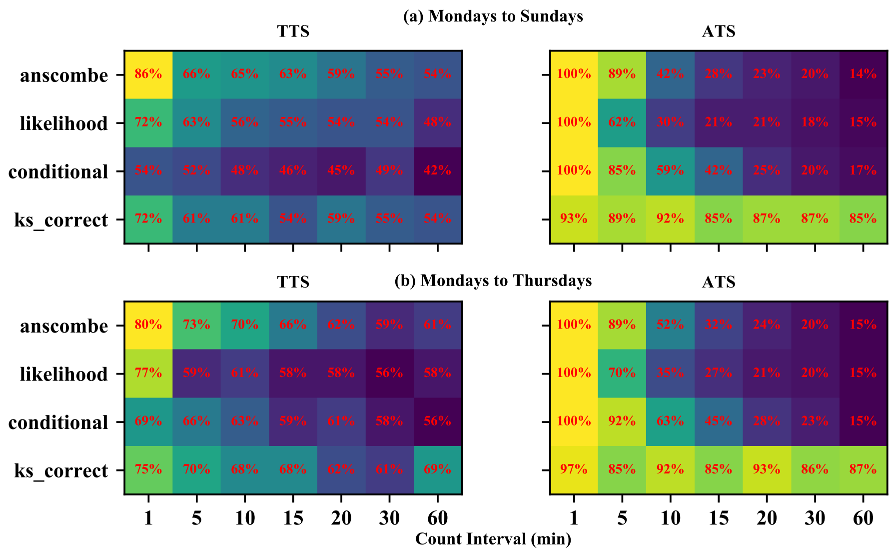


Fig. 3.20.: Hypothesis test for vehicle arrivals at Community Districts in 2-hour off peak

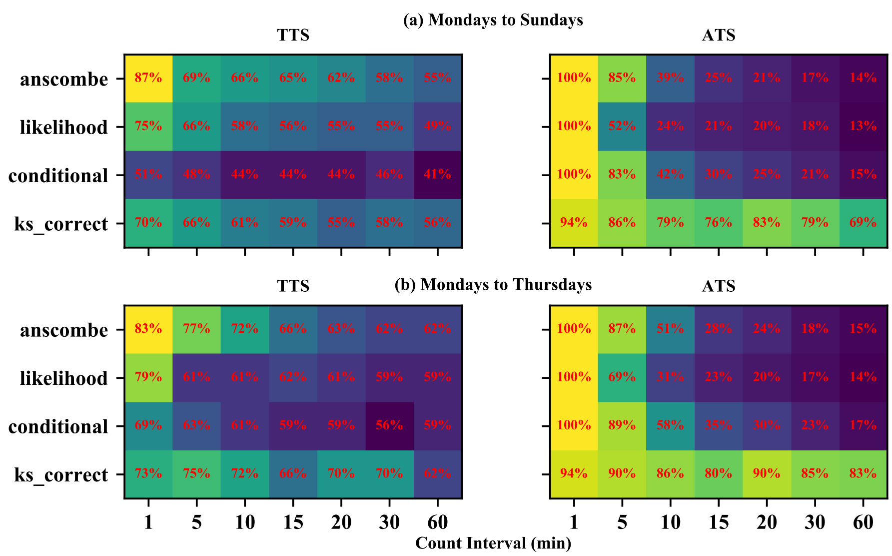
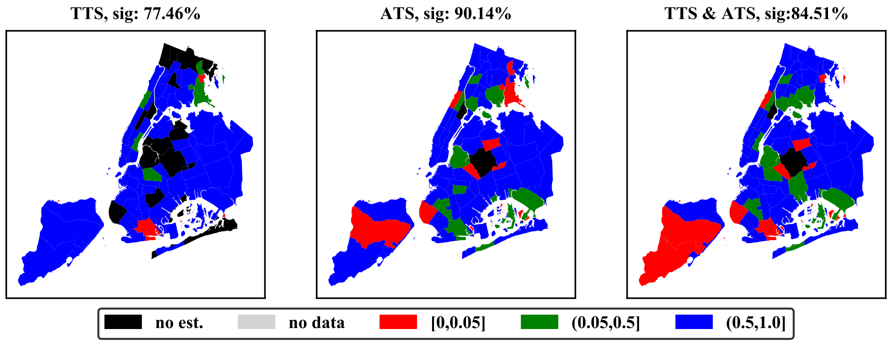
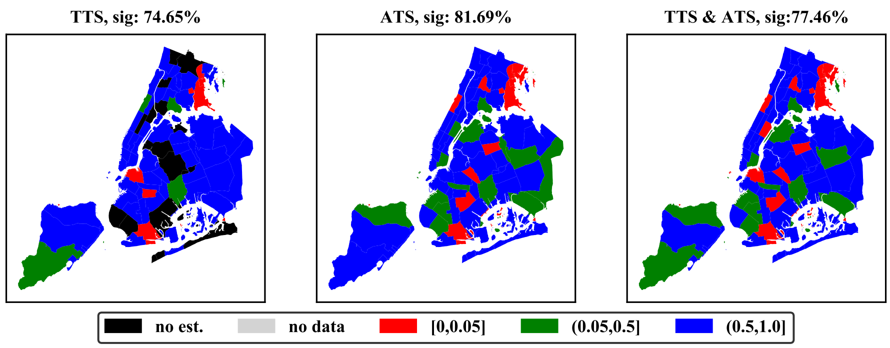


Fig. 3.21.: Hypothesis test for vehicle arrivals at Community Districts in 3-hour off peak

rare TTS and ATS activities. In NYC, most taxi activities concentrate in Manhattan downtown and midtown, as well as two airports, Brooklyn downtown, and



(a) 1-min count interval in 1-hour off peak



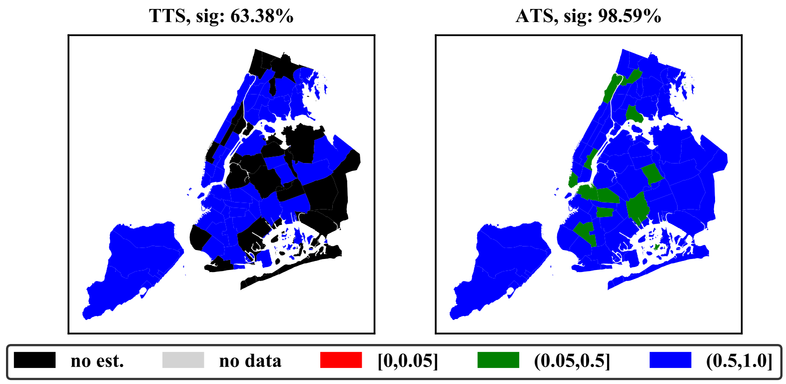
(b) 1-min count interval in 1-hour peak

Fig. 3.22.: Hypothesis test for passenger pickups by Community Districts in weekdays
Note: 'sig' indicates percentage of community districts not rejecting Poisson distribution, represented by blue and green color.

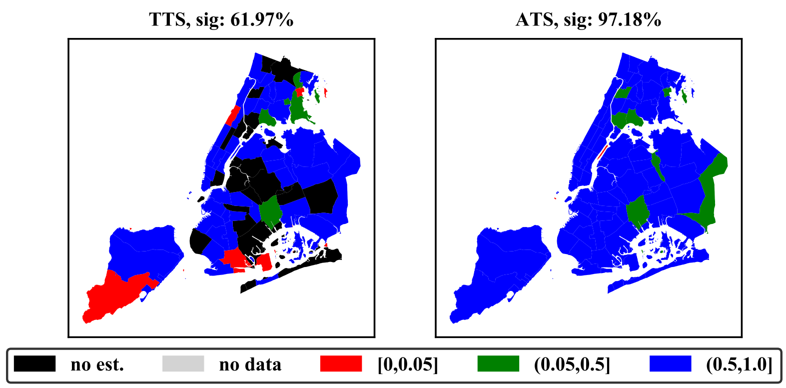
Queens downtown. Our hypothesis tests strongly support the Poisson assumption on passenger and vehicle arrivals in those areas. For more details on TTS and ATS activities and facts in NYC, you can refer to 2018 NYC TAXI FACT BOOK (http://www.nyc.gov/html/tlc/downloads/pdf/2018_tlc_factbook.pdf).

3.5 Summary

This chapter describes the detail procedures of large scale mobility dataset acquisition and processing, as well as determining spatial and temporal aggregations. We



(a) 1-min count interval in 1-hour off peak



(b) 1-min count interval in 1-hour peak

Fig. 3.23.: Hypothesis test for vehicle arrivals by Community Districts in weekdays

Note: 'sig' indicates percentage of community districts not rejecting Poisson distribution, represented by blue and green color.

first briefly introduce the data collection schemes on Uber platform. Then, we address imperfect and missing information through proposing ride estimation methods and existence of phantom vehicles. Moreover, our data preprocessing methods are validated with officially reported daily Uber ridership in NYC. Last, we select community district and 1-min time interval as spatial and temporal aggregation scale, regardless of peak and off peak hours, since there exists Poisson distribution of vehicle and passenger generations. Moreover, we fix our study period to 1-hour peak (6pm to 7pm) and 1-hour off peak (10am to 11am) on Mondays to Thursdays.

4. INFLUENCING FACTORS AND HETEROGENEITY IN DEMAND AND SUPPLY OF TRADITIONAL AND APP BASED TAXI SERVICES

4.1 Introduction

Taxis, both the TTS and ATS, are important urban transportation modes, which provide convenient 24/7 door-to-door mobility service for urban residents. For example, almost 10% of intra-city transit trips (i.e. bus, metro, and taxi trips) are by a fleet of more than 13,000 traditional yellow taxicabs in New York City (NYC), one of the largest taxi systems in North America. However, the market share of the traditional taxi system has been disrupted due to significant shifts in the taxi market. First, the NYC taxi and limousine commission introduced a feeder fleet of more than 7,000 green taxicabs in 2013, recognizing the unmet passengers in remote areas, for instance in region II as shown in Figure 1. Second, the app-based taxi services have exploded over the last seven years due to various factors including growth of smart phone usage. The new services, for instance, Uber, Lyft, and DIDI, provide diverse and competitive taxi services, and have reshaped this industry. In NYC, Uber has taken millions of Manhattan riders away from traditional yellow taxicabs in 2014 and 2015, which results in an average decrease of 50,000 in daily ridership by yellow taxicabs [44]. It has also seen similar or even worse decreases in other U.S. major cities, such as, 30% in Austin [150], 22% in Boston [151], and the bankruptcy of San Francisco's largest taxi firm in January 2016. All these changes are not uniform within a city and are driven by diverse factors. But what drives this heterogeneity and what are the influencing factors? This is the key question for this study to understand the influencing factors of the taxi ridership by both the traditional and app-based taxicabs.

Recent years have seen increasing discussions on taxi ridership driven by the availability of individual taxi trip records from GPS-equipped taxicabs. The literature mainly employed a generalized regression model for the traditional taxi ridership and explored primarily with exogenous impacts [29, 31, 32, 59]. As the rise of ATS, the users' acceptance were investigated mainly based on stated preference method [4, 43]. However, the literature is limited to one unique taxi ridership without any considerations of competition or impacts from other similar taxi services. A comprehensive modeling framework is needed for competing taxi services, considering the interactions of both TTS and ATS ridership. In particular, we have not seen extensive discussions on the impacts of dynamic pricing. Furthermore, the existing generalized regression cannot properly address the significance of spatiotemporal autocorrelations and heterogeneity. The spatial and temporal imbalance of taxi activities challenges a common assumption in generalized regression of the normal distribution of dependent variables. Recent studies have shown the efforts of employing spatial econometric analyses for spatial autocorrelations in taxi rides [32] and implementing independent analyses on hourly taxi rides [29]. However, we have not seen a perfect modeling structure or model estimation method for both spatiotemporal autocorrelations and heterogeneity. In addition to number of rides, there emerges multiple explorations on ATS supply generation, but with simple linear regression and without considerations over complex spatiotemporal structure [6].

This study fills the aforementioned research gaps and introduces appropriate modeling structures for demand and supply of both ATS and TTS. Starting from Poisson regression, we first examine spatial and modal correlations in residuals. Then we compare performance of spatial econometric model and geographically weighted regression model for spatial autocorrelation, as well as simultaneous equation system for modal correlation. Two case studies are developed based on traditional (i.e. yellow) taxi ride, app-based (i.e. Uber) taxi ride, and app-based taxi supply, observed in April, 2017 in NYC, along with other socioeconomic, transportation built environment, weather, dynamic pricing, shared bike system, and land use variables. To clarify the spatiotemporal heterogeneity, we also classify data samples into six groups depending on city boroughs (Manhattan, outer, or airports) and time periods (peak or off peak). Every group is recommended with one appropriate modeling structure. To our knowledge, this is one of the first studies to model such complex data with spatial, temporal and modal concerns.

The remaining sections are organized as follows: section 4.2 describes the data preparation; section 4.3 presents the representative modeling structures and estimation methods; section 4.4 explores the model performance and selection; section 4.5 and 4.6 summarize empirical findings on demand and supply generation, respectively; and section 4.7 discusses the major efforts, contributions, and future study.

4.2 Dependent and Explanatory Variables

The dependent variables of ridership, as well as Uber vehicle onlines, every minute are derived based on the spatial and temporal aggregations of corresponding ride and vehicle online events. First, we aggregate all pickup or vehicle online geolocations to 61 Community Districts (Note that about 10 Community Districts [CDs] are removed due to no green space), as shown in Figure 4.1. Second, we count events in every same minute and every Community districts as dependent variables. The observations are distributed in 9 weekdays of April, 2017. In final, we obtain 6,480 (i.e. 12 CDs $\times$ 9 days $\times$ 60 minutes), 25,380 (i.e. 47 CDs $\times$ 9 days $\times$ 60 minutes), and 1,080 (i.e. 2 CDs $\times$ 9 days $\times$ 60 minutes) observations during peak hours (18:00-19:00) for Manhattan, outer Manhattan, and airport, respectively. Meanwhile, we have another same sample size during off peaks (10:00-11:00).

Four groups of explanatory variables are initially extracted from multiple open data sets, including socioeconomic, land use, transit, and others. The first group of explanatory variables contains 5 variables related to demographics, education, income, vehicle ownership, commuting, and employment. The second group of explanatory variables contains 3 variables related to building floor area, commercial area, and

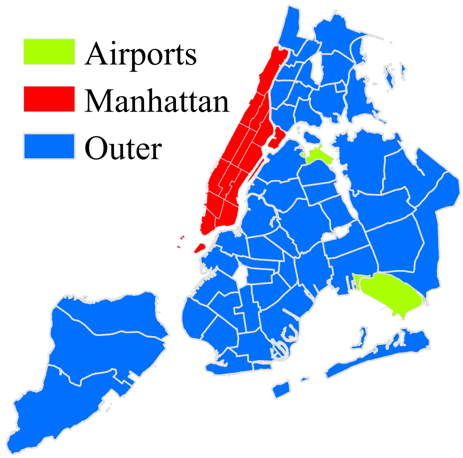


Fig. 4.1.: Study areas of interest

residential area. The third group of explanatory variables contains 3 variables related to metro stations, city bus stations, and bike facility. The last group of explanatory variables contains 3 variables related to weather, surge pricing multiplier, and shared bike usage. The definitions of all explanatory variables, as well as corresponding summary statistics, are presented in Table 4.1, 4.2, and 4.3. Moreover, the separate analyses by regions is recommended in another similar study [63].

One of multi-collinearity test methods are employed to filter out independent variables, which is designed to exclude those variables with high correlation coefficients. The step is necessary since most econometric models require independence of explanatory variables. The test yields one value of variance inflation factor (VIF) for every explanatory variable, which is the inverse of difference between 1 and goodness of fit R^2 value obtained by regressing the explanatory variable on others. In general, the rule of thumb indicates that keeping only explanatory variables with VIFs less than 10 can reduce the dependency. The corresponding results are also presented in above Tables.

Table 4.1.: Summary statistics for ridership and explanatory variables in Manhattan, NYC

Variables	Unit	Min	Max	Mean	Variation	VIF
Ridership/Supply						
Yellow Taxi -peak	per minute	0	128	26.75	27.34	n/a
Yellow Taxi -off peak	per minute	0	96	18.72	19.46	n/a
Uber -peak	per minute	0	65	12.37	10.75	n/a
Uber -off peak	per minute	0	30	6.37	4.70	n/a
Yellow and Uber -peak	per minute	0	177	39.12	36.75	n/a
Yellow and Uber -off peak	per minute	0	113	25.09	23.28	n/a
Uber onlines -peak	per minute	0	9	1.01	1.14	n/a
Uber onlines -off peak	per minute	0	7	0.60	0.84	n/a
Socioeconomics						
Ratio of workers not owning any vehicles	n/a	0.67	0.81	0.73	0.04	3.48
Ratio of residents with Bachelor's degree or higher	n/a	0.25	0.71	0.47	0.16	>10
Employment ratio over total residents	n/a	0.50	0.76	0.64	0.08	4.77
Median household annual income	per 5,000$	7.55	36.11	18.51	8.63	>10
Mean commuting time	minutes	22.47	40.58	30.56	5.10	8.52
Land use[1]						
Development intensity	n/a	1.32	12.15	4.54	2.83	>10
Commercial intensity	n/a	0.42	10.85	2.47	2.82	6.49
Residential intensity	n/a	0.90	4.06	2.08	0.94	5.00
Transportation Built Environment[2]						
Bike rack intensity	per mi^2	15.32	214.29	86.26	62.06	4.10
Metro station intensity	per mi^2	0.81	7.37	2.83	2.07	4.86
CityBus station intensity	per mi^2	21.94	50.40	37.28	8.64	2.73
Other						
Rain, 1 if has non-zero precipitation in that hour, 0 if otherwise	per n/a	0	1	0.88 vs.	0.12	1.02
Surge Pricing Multiplier -peak	n/a	1.00	1.84	1.05	0.13	1.15
Surge Pricing Multiplier -off peak	n/a	1.00	1.00	1.00	0	n/a
CitiBike[3] ridership -peak	per minute	0	43	6.56	6.98	4.50
CitiBike[3] ridership -off peak	per minute	0	18	2.22	2.54	2.43

1. Intensity for land use indicates the ratio of corresponding building area to region area;
2. Intensity for transportation facilities indicates number of facilities over region area;
3. CitiBike is the shared bike system in NYC.

Table 4.2.: Summary statistics for ridership and explanatory variables in outer Manhattan, NYC

Variables	Unit	Min	Max	Mean	Variation	VIF
Ridership/Supply						
Yellow Taxi -peak	per minute	0	10	0.12	0.49	n/a
Yellow Taxi -off peak	per minute	0	4	0.07	0.32	n/a
Uber -peak	per minute	0	18	2.13	2.13	n/a
Uber -off peak	per minute	0	15	1.54	1.58	n/a
Yellow and Uber -peak	per minute	0	27	2.25	2.36	n/a
Yellow and Uber -off peak	per minute	0	15	1.61	1.67	n/a
Uber onlines -peak	per minute	0	7	0.40	0.68	n/a
Uber onlines -off peak	per minute	0	6	0.31	0.60	n/a
Socioeconomics						
Ratio of workers not owning any vehicles	n/a	0.03	0.67	0.41	0.18	5.97
Ratio of residents with Bachelor's degree or higher	n/a	0.07	0.61	0.22	0.11	4.16
Employment ratio over total residents	n/a	0.45	0.71	0.57	0.06	2.87
Median household annual income	per 5,000$	4.56	22.60	11.07	3.87	>10
Mean commuting time	minutes	32.95	49.82	42.57	3.44	2.19
Land use[1]						
Development intensity	n/a	0.21	2.05	1.00	0.48	>10
Commercial intensity	n/a	0.02	1.10	0.27	0.19	5.90
Residential intensity	n/a	0	1.70	0.73	0.39	2.21
Transportation Built Environment[2]						
Bike rack intensity	per mi^2	0.60	102.10	18.85	26.22	2.76
Metro station intensity	per mi^2	0	2.58	0.83	0.64	3.80
CityBus station intensity	per mi^2	7.24	40.23	23.82	7.62	2.76
Other						
Rain, 1 if has non-zero precipitation in that hour, 0 if otherwise	per n/a	0	1	0.88 vs.	0.12	1.00
Surge Pricing Multiplier -peak	n/a	1.00	2.00	1.00	0.04	1.03
Surge Pricing Multiplier -off peak	n/a	1.00	1.71	1.00	0.03	1.01
CitiBike[3] ridership -peak	per minute	0	15	0.26	1.11	2.50
CitiBike[3] ridership -off peak	per minute	0	8	0.11	0.53	1.68

1. Intensity for land use indicates the ratio of corresponding building area to region area;
2. Intensity for transportation facilities indicates number of facilities over region area;
3. CitiBike is the shared bike system in NYC.

Table 4.3.: Summary statistics for ridership and explanatory variables at Airports

Variables	**Unit**	**Min**	**Max**	**Mean**	**Variation**	**VIF**
Ridership/Supply						
Yellow Taxi -peak	per minute	0	26	9.14	3.70	n/a
Yellow Taxi -off peak	per minute	0	24	7.74	4.53	n/a
Uber -peak	per minute	0	19	4.97	2.81	n/a
Uber -off peak	per minute	0	18	4.12	2.87	n/a
Yellow and Uber -peak	per minute	3	37	14.01	5.14	n/a
Yellow and Uber -off peak	per minute	1	34	11.86	6.52	n/a
Uber onlines -peak	per minute	0	6	0.83	0.98	n/a
Uber onlines -off peak	per minute	0	5	0.59	0.84	n/a
Land use[1]						
Commercial intensity	n/a	0.05	0.08	0.07	0.01	>10
Transportation Built Environment[2]						
Metro station intensity	per mi^2	0	0.06	0.03	0.03	>10
CityBus station intensity	per mi^2	3.2	10.86	7.03	3.83	1
Other						
Rain, 1 if has non-zero precipitation in that hour, 0 if otherwise	per n/a	0	1	0.88 vs.	0.12	1
Surge Pricing Multiplier -peak	n/a	1	1	1	1	n/a
Surge Pricing Multiplier -off peak	n/a	1	1.6	1.01	0.06	1.02

1. Intensity for land use indicates the ratio of corresponding building area to region area;
2. Intensity for transportation facilities indicates number of facilities over region area;
3. CitiBike is the shared bike system in NYC.

4.3 Methodology

4.3.1 Spatial Lag Poisson model

One of representative ways to relax spatial dependency is the spatial lag model extending general forms of linear regression with a weighted component, as shown in equation4.1. Since our study involves count data structure, we further adapt the generalized form of spatial lag model with Poisson regression as shown in equation 4.2 and 4.3. The major step is to integrate weighted dependent variable into Poisson regression as one independent variable.

$$Y = \rho WY + X\beta + \varepsilon \tag{4.1}$$

$$\mu = \hat{y} + X\beta + \varepsilon \tag{4.2}$$

$$Pr(y) = \frac{e^{-\mu}\mu^{y}}{y!} \tag{4.3}$$

where, Y is the vector of rides or supply; W is a $n \times n$ weight matrix, described later; ρ is the spatial lag to be estimated; X is a $n \times k$ matrix consisting of explanatory variables; β is a vector with estimated coefficients for k explanatory variables; ε is the error term; n is the number of spatial units; k is the number of explanatory variables; y is an observation on ride or supply; μ is the Poisson arrival rate; and $\hat{y}$ is the weighted rides or supply from neighboring spatial units, also the element of WY.

We also introduce a Gaussian kernel function based method to determine weight matrix W, as shown in equation 4.4. The Gaussian kernel function is one of distance decay functions, which has been successfully applied in a similar study on taxi ridership [32]. The ride or supply in one spatial unit will contribute to that in another neighboring one whose distance is less than bandwidth b from the centroid. However, the contribution will decrease as the distance of two regions increases. If two spatial units are over bandwidth b, both rides or supply generations are independent in both spatial units.

$$w_{ij} = \begin{cases} exp[-\frac{1}{2}(\frac{d_{ij}}{b_i})^2] & d_{ij} < b_i \\ 0 & otherwise \end{cases} \tag{4.4}$$

where, w_{ij} is the weight for spatial unit pair of i and j; d_{ij} is the spatial distance between two spatial units i and j; and b is the bandwidth, determined based on model performance.

4.3.2 Geographically Weighted Poisson Regression

The spatial lag model is deigned for spatial autocorrelation, but not for spatial non-stationary condition (i.e. impacts vary across spatial units). For the latter one, we introduce Geographically Weighted Poisson Regression (GWPR) as equation 4.5. The GWPR is the extension of Geographically Weighted Regression model for count data, which can also address spatial non-stationary and estimate local coefficients. The model estimation performs at each spatial unit through introducing a weighting function thus can include influencing factors from neighboring spatial units, as shown in equation 4.6 and 4.7. The weighting function is also Gaussian kernel, shown in equation 4.4.

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} x_{itk} + \gamma_i t + \delta_{it}(p_{it} \times D_i) + \varepsilon_{it} \quad \mathbf{(4.5)}$$

$$\widehat{\beta}_i = (XW_iX)^{-1} X^T W_i Y \quad \mathbf{(4.6)}$$

$$W_i = diag(w_{i1}, w_{i2}, \cdots, w_{ij}, \cdots, w_{in}) \quad \mathbf{(4.7)}$$

where, Y_{it} is the ridership or entering drivers at spatial unit i in time of t; β_{ik} is the local coefficient for variable k at location i; and n is the number of spatial units.

4.3.3 Estimation Methods

For the spatial Lag Poison model, a two-stage estimation approach is developed, including determining weight matrix and weighted dependent variables, then estimation of Poisson regression with 'glm' package in R. In contrast, the GWPR model requires additional steps due to local estimations. We refer to an iterative approach that can yield time-invariant local coefficients.

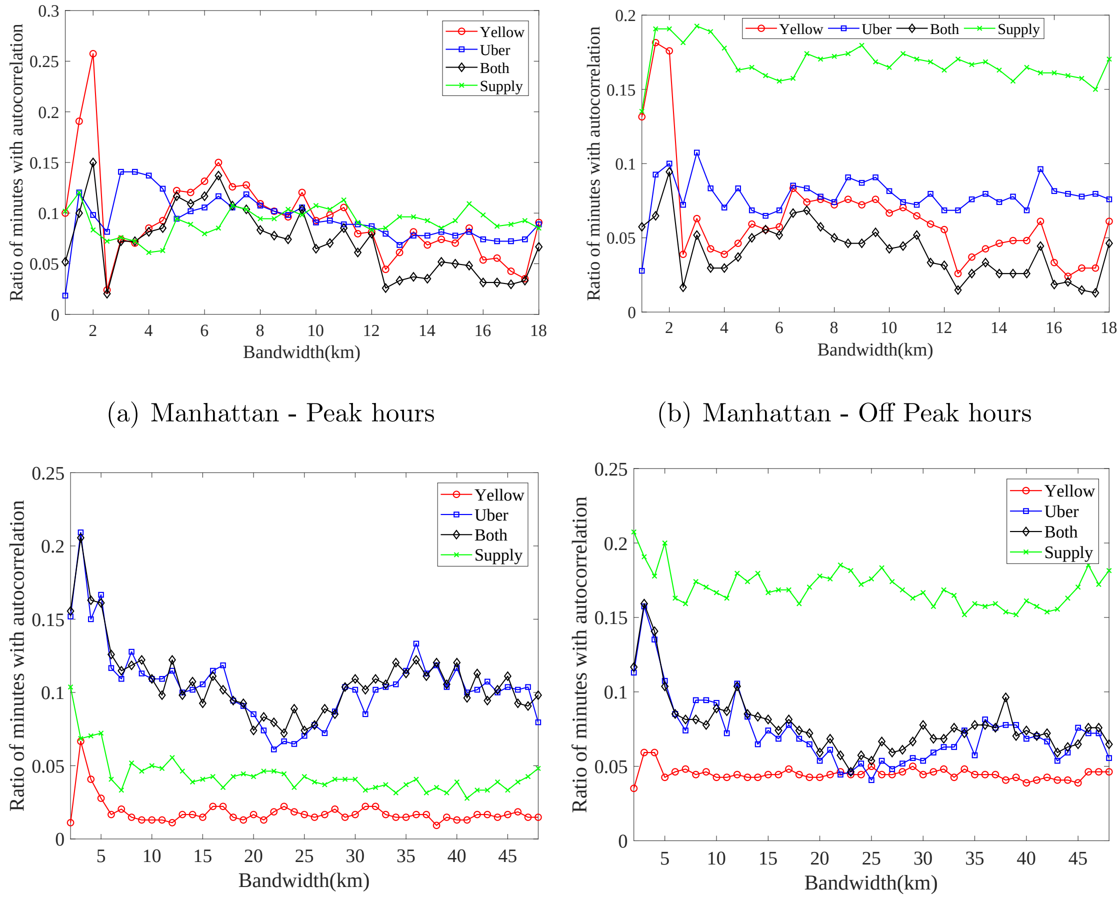
1. Select a local bandwidth b_i and generate weighting matrix W_i at spatial unit i;
2. Subsample observed data for local estimation at spatial unit i;
3. Iterate over time and Weight all explanatory variables with weighting matrix W_i;
4. Apply Poisson regression to the weighted subsample data;
5. Repeat 2 to 4, till all spatial units are estimated.

4.4 Model Settings and Performance

4.4.1 Spatial and Modal Autocorrelation in Poisson Regression

We simply apply Poisson regression for yellow taxi ridership, Uber ridership, the both services ridership, as well as Uber supply (in terms of Uber online vehicles), in each combination scenario of Manhattan or outer Manhattan, and peak or off peak. Although the goodness of fit indicates good specifications for ridership in Manhattan, we are more interested in whether the model accounts for potential spatial and modal autocorrelations in this panel dataset.

Regarding the spatial autocorrelation test, we introduce Global Moran's I test, as well as distance based weight matrix [63]. The null hypothesis of Global Moran's I test states that the ridership or supply is randomly distributed among Community Districts in Manhattan or outer Manhattan. In every minute (540 in total), we run Global Moran's I test for Poisson regression residuals, given more than 30 levels of weight matrix. Then we count ratio of minutes with significant spatial autocorrelation under confidence level of 95%. Figure 4.2 summarizes ratios in four different scenarios. Overall, the ratios are at very low level, less than 0.3. In other words, only in less than 30% of minutes, the spatial autocorrelation is one major concern. In specifics, the ridership, regardless of yellow taxi, Uber, and both services, has similar spatial autocorrelation patterns, increasing at beginning and decreasing as



(a) Manhattan - Peak hours

(b) Manhattan - Off Peak hours

(c) Outer Manhattan - Peak hours

(d) Outer Manhattan - Off Peak hours

Fig. 4.2.: Results of Spatial Autocorrelation Tests in each minute

bandwidth increases. The adaptive bandwidths (i.e. the bandwidth with maximum ratio) are with very small variations, from minimum of 1.5 km to maximum of 3 km. However, the spatial autocorrelation patterns are different for Uber supply between peak and off peak hours. The off peak hours exhibit higher probability of significant spatial autocorrelation. Considering the existence of spatial autocorrelation, though small probability, we still introduce spatial lag model and GWPR to empirically validate whether addressing spatial autocorrelation could improve demand and supply specifications.

Table 4.4.: Pearson Correlation Tests for Poisson regression residuals

Location	Period	t-statistic	degree of freedom	p-val	correlation coefficient
Manhattan	Peak hours	11.53	6478	≈ 0	0.14
Manhattan	Off peak hours	8.6	6478	≈ 0	0.11
Outer Manhattan	Peak hours	1.94	25378	0.06	0.01
Outer Manhattan	Off peak hours	1.65	25378	0.1	0.01
Airports	Peak hours	4.81	1078	≈ 0	0.14
Airports	Off peak hours	3.49	1078	≈ 0	0.11

Regarding the modal autocorrelation test, we employ Pearson correlation test. The null hypothesis of the test is that the true correlation is a specific value zero and the alternative hypothesis indicates that true correlation value is not equal to zero. In each region of Manhattan, outer Manhattan, and airports, we implement Pearson correlation test on all residuals from Poisson regression for yellow taxi and Uber rides in 540 minutes. The test results are summarized in Table 4.4. It is apparent that only outer Manhattan region has no statistically significant correlation between yellow taxi and Uber rides. But the correlation coefficients are very small in other two regions, less than 0.15. Considering the existence of modal correlations in Manhattan and airports, we also compare multiple estimation strategies for yellow taxi and Uber demand, including an overall demand for both yellow and Uber services, separate demand for each service but combine with simultaneous equation system, and separate demand estimation.

4.4.2 Model Selections

We select multiple performance metrics, including log-likelihood, goodness of fit R^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). In addition, we employ likelihood ratio test for conditions that two models have similar performance metrics. The test is based on χ^2 distribution and the statistic is computed from two times of differences in log-likelihood values. The null hypothesis is that introducing additional parameters can improve model performance.

Table 4.5 summarizes performance metrics of models based on Manhattan study cases. First, the simultaneous equation system cannot account for modal correlations very well. Separate analyses for yellow and Uber rides lead to a slight better fitting. Alternatively, estimating overall rides rather than each service can also have better fitting. Second, addressing spatial autocorrelations can significantly improve model performance for both demand and supply. Both spatial lag Poisson and GWPR models reduce the probability of spatial autocorrelation and the latter model has a larger reductions in particular during peak hours. This indicates complex spatiotemporal heterogeneity in taxi system that demand and supply are more likely to be spatial non-stationary during peak hours but spatial stationary during off peak hours. In addition, the both models are comparative in model fitting. However, the likelihood ratio test shows that GWPR cannot significantly outperform spatial lag Poisson model, due to much more estimators.

Table 4.6 exhibits performance metrics of models based on outer Manhattan study cases. Same as the Manhattan cases, the simultaneous equation system does not perform well, compared to separate analyses for yellow and Uber rides. Moreover, estimating overall rides does not have better fitting. This is because there are no statistical evidence of correlated yellow taxi and Uber demand in outer Manhattan. Second, addressing spatial autocorrelations can significantly improve model performance for both demand and supply. Both spatial lag Poisson and GWPR models reduce the probability of spatial autocorrelation. However, the demand and supply in outer Manhattan behave more identically across spatial units. The GWPR does not yield a good fitting, due to the assumption of spatial non-stationary states.

Table 4.7 shows performance metrics of models based on airport study cases. Same as aforementioned two regions, the simultaneous equation system does not perform well, compared to separate analyses for yellow and Uber rides. Alternatively, estimating overall rides rather than each service can also have better fitting, because of

Table 4.5.: Model performance in Manhattan

Performance Metric	Poisson Regression (yellow, Uber, both, Uber supply)	Simultaneous Equation System (yellow & Uber)	Spatial Lag (yellow, Uber, both, Uber supply)	GWPR (yellow, Uber, both, Uber supply)
peak hours (sample size n=6480)				
AIC	50027,34266,52999,16878	87422	36884,30873,40270,12681	35109,33529,41140,16425
BIC	50048,34287,53017,16892	87464	36907,30895,40292,12691	35588,33940,41595,16611
Log-likelihood	-25001,-17120,-26489,-8431	-43691	-18429,-15424,-20123,-6334	-17519,-16718,-20546,-8191
Degree of Freedom	12,12,10,8	20	13,12,12,6	144,132,132,60
R^2	0.87,0.84,0.90,0.11	0.82 (0.81&0.82)	0.94,0.90,0.96,0.58	0.95,0.86,0.96,0.16
Ratio of minutes with autocorrelation	0.25,0.14,0.15,0.12	n/a	0.21,0.06,0.11,0.07	0.05,0.06,0.05,0.06
off peak hours (sample size n=6480)				
AIC	46504,29554,49832,13240	75595	32967,26333,35561,10269	31548,28722,36465,13218
BIC	46524,29573,49851,13254	76646	32988,26351,35581,10277	31985,29090,36857,13377
Log-likelihood	-23241,-14765,-24904,-6611	-38273	-16471,-13156,-17769,-5129	-15741,-14339,-18199,-6597
Degree of Freedom	11,11,11,8	24	12,10,11,5	132,108,120,48
R^2	0.83,0.66,0.85,0.05	0.81 (0.81&0.66)	0.93,0.78,0.85,0.45	0.95,0.69,0.94,0.06
Ratio of minutes with autocorrelation	0.18,0.11,0.09,0.19	n/a	0.21,0.04,0.09,0.14	0.16,0.13,0.14,0.18

Table 4.6.: Model performance in outer Manhattan

Performance Metric	Poisson Regression (yellow, Uber, both, Uber supply)	Simultaneous Equation System (yellow & Uber)	Spatial Lag (yellow, Uber, both, Uber supply)	GWPR (yellow, Uber, both, Uber supply)
peak hours (sample size n=25380)				
AIC	11261,87474,88614,40693	123369	9435,73962,76239,27932	10659,81815,82682,38995
BIC	11287,87505,88645,40721	123420	9466,73992,76270,27956	12913,84141,84936,40712
Log-likelihood	-5619,-43724,-44293,-20334	-61665	-4704,-36967,-38106,-13956	-5214,-40828,-41226,-19424
Degree of Freedom	11,13,13,12	19	13,13,13,10	564,564,564,423
R^2	0.54,0.34,0.39,0.06	0.4 (0.24&0.41)	0.66,0.60,0.61,0.57	0.59,0.45,0.50,0.14
Ratio of minutes with autocorrelation	0.07,0.21,0.21,0.10	n/a	0.10,0.01,0.03,0.06	0.41,0.06,0.08,0.05
off peak hours (sample size n=25380)				
AIC	8241,78019,78749,35796	94721	6427,64337,65268,23451	7910,73371,74082,34589
BIC	8268,78048,78777,35813	94781	6451,64366,65299,23475	9585,75494,76412,36358
Log-likelihood	-4109,-38997,-39362,-17889	-47338	-3203,-32156,-32621,-11715	-3861,-36608,-36964,-17247
Degree of Freedom	11,12,12,8	22	10,12,13,10	423,517,564,423
R^2	0.49,0.22,0.26,0.04	0.27 (0.21&0.27)	0.67,0.55,0.57,0.58	0.54,0.34,0.37,0.10
Ratio of minutes with autocorrelation	0.06,0.16,0.16,0.21	n/a	0.09,0.01,0.01,0.17	0.50,0.07,0.09,0.18

Table 4.7.: Model performance in airports

Performance Metric	Poisson Regression (yellow, Uber, both, Uber supply)	Simultaneous Equation System (yellow & Uber)
peak hours (sample size n=1080)		
AIC	5750,5025,6338,2611	10866
BIC	5753,5027,6338,2613	10876
Log-likelihood	-2872,-2510,-3166,-1303	-5425
Degree of Freedom	3,2,3,2	8
R^2	0.12,0.15,0.23,0.04	0.13(0.13&0.14)
off peak hours (sample size n=1080)		
AIC	5233,4661,5912,2215	10317
BIC	5237,4665,5916,2218	10328
Log-likelihood	-2612,-2326,-2952,-1104	-5150
Degree of Freedom	4,4,4,3	8
R^2	0.58,0.35,0.63,0.04	0.5 (0.57&0.32)

significant modal correlation. The time of day also influences the model performance. The off peak hours have better model performance, which is different from Manhattan and outer Manhattan cases. Moreover, the supply estimation at airports are also not very good, with much lower goodness of fit values. This is likely because there are very limited Uber vehicle online at airports.

To sum up, we should care much more about the spatial autocorrelation, while specifying demand and supply of ATS and TTS. Addressing spatial autocorrelation can greatly improve model performance. Although the yellow taxi demand in Manhattan reveals spatial non-stationary impacts, the spatial lag Poisson model performs well enough and GWPR model lacks statistical evidence of better performance. On the other hand, we find the weak correlation between yellow and Uber demand, primarily in Manhattan and at airports. One solution is to simply estimate the overall demand of both services. A simultaneous equation system with correlated error terms may perform even worse than separate estimations for each service. The recommended

Table 4.8.: Model selections for demand and supply estimation

Periods	Objects	Manhattan	Outer Manhattan	Airports
Peak hours	yellow taxi rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Uber rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Overall rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Uber supply	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
Off peak hours	yellow taxi rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Uber rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Overall rides	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression
	Uber supply	Spatial Lag Poisson	Spatial Lag Poisson	Poisson Regression

modeling structures are presented in Table 4.8. For most cases, we recommend spatial lag Poisson model in Manhattan and outer Manhattan, as well as Poisson regression at airports.

4.5 Empirical Findings on Demand

Table 4.9, 4.10, and 4.11 present parameter estimates and marginal effects for overall demand in Manhattan, yellow and Uber demand in outer Manhattan, and overall demand at airports, respectively. Overall, the sets of significant influencing factors are similar between Manhattan and outer Manhattan, regardless of peak and off peak hours.

Among all variables, we are more interested in the impacts of dynamic pricing on taxi demand. Unsurprisingly, there are negative impacts on overall taxi demand in Manhattan during peak hours. One unit increase in surge pricing multiplier will yield 1.82 reduction in overall taxi rides. During off peak, the dynamic pricing multiplier has no significant effects, likely due to less opportunities of applying higher surge pricing multipliers. Also in outer Manhattan, we can observe negative impacts of surge pricing multiplier on Uber rides, regardless of peak and off peak hours. However,

Table 4.9.: Estimation results for demand in Manhattan, NYC

Variables	**Peak hours**			**Off peak hours**		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	1.588	22.37	n/a	0.598	45.49	n/a
spatial lag	0.018	116.68	0.692	0.031	147.38	0.789
Socioeconomics						
Mean commuting time	-0.014	-6.08	-0.545	n/a	n/a	n/a
Land use						
Commercial intensity	0.018	5.91	0.691	0.016	8.99	0.391
Residential intensity	0.358	96.5	14.02	0.28	73.83	7.03
Transportation Built Environment						
Bike rack intensity	0.002	23.97	0.08	0.003	50.05	0.073
Metro station intensity	0.191	73.56	7.48	0.16	64.98	4.018
CityBus station intensity	-0.004	-6.02	-0.17	n/a	n/a	n/a
Other						
Surge Pricing Multiplier	-0.047	-3.74	-1.82	n/a	n/a	n/a

the marginal effects are much lower than those in Manhattan. This is because outer Manhattan residents lack alternative transportation modes and cannot transfer others if facing higher surge pricing multipliers. The airports have a different pattern, only observing negative impacts during off peak hours and surprisingly positive impacts.

Second, we also observe positive spatial lag parameters in all study cases. This indicates the spatial units with high (low) ridership are generally surrounded by ones also with high (low) ridership. More importantly, this spatial autocorrelation impacts do not vary temporally, given the facts that similar marginal effects between peak and off peak hours.

Third, the variables of land use and transportation built environment are common for demand estimation, but their impacts vary spatially and temporally. Both commercial and residential intensity has positive impacts on taxi demand, besides residential intensity for yellow taxi rides in outer Manhattan. However, the marginal effects indicates a relatively higher increase in taxi demand during peak hours and Manhattan. In particular, the one unit increase in residential intensity can lead to 14 additional rides in Manhattan during peak hours. On the other hand, transportation built environment presents not only different marginal effects but also different influences. In Manhattan, it is more likely to have rides in spatial units with high bike rack and metro intensity and low city bus station intensity. However, it does not hold for outer Manhattan cases of both yellow and Uber services, as well as airport cases. Regardless of negative and positive effects, most marginal effects are at very low level, less than 0.01 increases in demand, except for impacts of metro station intensity in Manhattan (more than 4).

In addition, we also identify the significant impacts of socioeconomics. Commuters with long travel times are less likely to hail a traditional taxicab but may prefer to Uber as alternative to other transportation modes. Remaining socioeconomic variables are only significant in outer Manhattan, including workers without available vehicles, residents with Bachelor's degree or higher, and employment rate. Unsurprisingly, all these population related variables positively contribute to both yellow taxi and Uber demand. In particular the employment ratio, one unit increase will lead to more than 1 yellow taxi ride and 3 Uber rides.

In final, the weather of raining will increase overall taxi rides by 1.35, from airports during peak hours. Correspondingly, it decreases outer Manhattan yellow taxi rides by 0.016. Although airport taxi rides may reduce during raining time, we do not identify significant changes in taxi demand in other regions. Moreover, shared bike system usage may also negatively influence Uber rides in outer Manhattan, possibly due to mode competitions.

Table 4.10.: Estimation results for demand in outer Manhattan, NYC

Variables	Peak hours-Yellow Taxi			Off peak hours-Yellow Taxi			Peak hours-Uber			Off peak hours-Uber		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	-7.824	-8.8	n/a	-14.09	-13.187	n/a	-1.191	-7.74	n/a	-3.5	-18.07	n/a
spatial lag	0.6	48.91	0.065	1.227	46.63	0.084	0.249	127.82	0.526	0.376	131.42	0.578
Socioeconomics												
Ratio of workers not owning any vehicles	4.997	13.59	0.542	1.759	4.71	0.12	0.417	7.14	0.88	n/a	n/a	n/a
Ratio of residents with Bachelor's degree or higher	3.586	7.03	0.389	n/a	n/a	n/a	0.826	10.05	1.743	0.45	5.19	0.692
Employment ratio over total residents	10.703	9.47	1.16	17.962	17.69	1.227	1.675	12.19	3.533	2.146	13.52	3.301
Mean commuting time	-0.075	-6.07	-0.008	-0.045	-3.03	-0.003	0.02	10.37	0.043	0.02	9.73	0.031
Land use												
Commercial intensity	2.427	8	0.263	2.834	8.57	0.194	0.281	6.3	0.593	0.189	3.79	0.291
Residential intensity	-0.613	-3.28	-0.066	n/a	n/a	n/a	0.123	5.63	0.26	0.173	7.4	0.267
Transportation Built Environment												
Bike rack intensity	-0.014	-13.13	-0.002	-0.01	-8.07	-0.001	-0.004	-16.34	-0.009	-0.005	-16.3	-0.007
Metro station intensity	-1.007	-8.53	-0.109	-0.775	-5.38	-0.053	-0.038	-2.77	-0.08	0.057	3.76	0.088
CityBus station intensity	0.053	7.05	0.006	0.04	5.38	0.003	0.018	17.97	0.039	0.018	14.78	0.027
Other												
Rain	-0.148	-2.06	-0.016	-0.217	-2.95	-0.015	n/a	n/a	n/a	n/a	n/a	n/a
Surge Pricing Multiplier	-1.995	-9.09	-0.216	n/a	n/a	n/a	-1.443	-15.39	-3.043	0.416	2.9	0.639
CitiBike ridership	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	-5.36	-0.042	-0.049	-6.37	-0.076
Surge Pricing Multiplier (1-min lag)	-1.449	-2.69	-0.157	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Surge Pricing Multiplier (5-min lag)	0.934	2.66	0.102	n/a	n/a	n/a	-0.68	-4.68	-1.431	n/a	n/a	n/a

Table 4.11.: Estimation results for demand at airports, NYC

Variables	**Peak hours**			**Off peak hours**		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	2.296	122.78	n/a	0.934	8.76	n/a
Transportation Built Environment						
CityBus station intensity	0.045	21	0.635	0.116	45.35	1.379
Other						
Rain	0.096	3.86	1.349	-0.22	-7.14	-2.613
Surge Pricing Multiplier	n/a	n/a	n/a	0.64	6.12	7.596

4.6 Empirical Findings on Supply

Table 4.12, 4.13, and 4.14 present parameter estimates and marginal effects for Uber supply in all three regions, respectively. Overall, Manhattan and outer Manhattan share common significant variables, although outer Manhattan requires more explanatory variables to specify Uber supply. More importantly, we observe more variables with temporally heterogeneous impacts even within one same region.

Among all significant variables, we first examine impacts of dynamic pricing on Uber supply. Different from significant deductions in demand by high surge pricing multiplier, there are almost no significant impacts on Uber supply. In other words, Uber drivers are not responsive to higher surge pricing multiplier. Only in outer Manhattan, a negative impact is captured during peak hours. The finding is consistent with some public reports that Uber driver partners are not chasing surge pricing due to short time periods.

Second, we also observe positive spatial lag parameters in all study cases. This indicates the spatial units with high (low) Uber supply are generally surrounded by ones also with high (low) Uber supply. More importantly, this spatial autocorrelation impacts do not vary temporally, given the facts that similar marginal effects between peak and off peak hours.

Table 4.12.: Estimation results for supply in Manhattan, NYC

Variables	**Peak hours**			**Off peak hours**		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	-2.894	-13.08	n/a	-1.556	-4.7	n/a
spatial lag	0.57	75.76	0.573	1.235	63.67	0.741
Socioeconomics						
Ratio of workers not owning any vehicles	2.774	9.43	2.792	1.719	4.04	1.03
Mean commuting time	n/a	n/a	n/a	-0.039	-9.97	-0.023
Land use						
Commercial intensity	0.4	8.73	0.04	n/a	n/a	n/a
Residential intensity	0.131	8.68	0.132	n/a	n/a	n/a
Transportation Built Environment						
Bike rack intensity	n/a	n/a	n/a	-0.001	-3.32	-0.001
CityBus station intensity	-0.01	-5.52	-0.01	n/a	n/a	n/a

Table 4.13.: Estimation results for supply in outer Manhattan, NYC

Variables	**Peak hours**			**Off peak hours**		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	-2.814	-9.41	n/a	-3.048	-17.83	n/a
spatial lag	1.007	148.06	0.4	1.2	139.64	0.376
Socioeconomics						
Ratio of workers not owning any vehicles	-0.484	-5.22	-0.192	-0.807	-6.14	-0.253
Ratio of residents with Bachelor's degree or higher	-0.819	-4.66	-0.325	-1.086	-5.19	-0.34
Employment ratio over total residents	3.017	9.72	1.197	1.769	5.13	0.554
Land use						
Commercial intensity	n/a	n/a	n/a	0.419	3.37	0.131
Residential intensity	0.32	6.85	0.127	0.487	8.78	0.152
Transportation Built Environment						
Bike rack intensity	-0.001	-2.17	-0.0005	0.002	3.4	0.001
Metro station intensity	n/a	n/a	n/a	-0.181	-5.2	-0.057
CityBus station intensity	0.01	4.18	0.004	0.01	3.66	0.003
Other						
Surge Pricing Multiplier	-0.71	-2.7	-0.282	n/a	n/a	n/a
CitiBike ridership	0.031	3.02	0.012	n/a	n/a	n/a
Surge Pricing Multiplier (3-min lag)	0.872	1.97	0.344	-0.893	-1.95	-0.278

Table 4.14.: Estimation results for supply ar airports, NYC

Variables	**Peak hours**			**Off peak hours**		
	Para.	**Statistic**	**Marginal**	**Para.**	**Statistic**	**Marginal**
constant	-0.692	-8.71	n/a	-1.017	-10.68	n/a
Transportation Built Environment						
CityBus station intensity	0.067	7.44	0.056	0.069	6.45	0.041
Other						
Rain	n/a	n/a	n/a	-0.283	-2.01	-0.168

Third, the variables of land use and transportation built environment are almost same for supply specifications, except for metro station intensity. Both commercial and residential intensity positively contribute to Uber supply with almost same marginal effects of around 0.13. However, we can not capture such impacts in Manhattan during off peak hours. This is likely because most Uber driver partners are part-time service providers and only work before or after work. During off peak hours, it is less likely to capture many online vehicles from those regions. On the other hand, transportation built environment presents negative impacts. Uber drivers tend to avoid regions with intense bike racks, metro stations, and city bus stations. However, there is also one special case in outer Manhattan that Uber drivers are more likely to become online in spatial units with more city bus stations. It may be determined by the advantages of Uber services over city bus in outer Manhattan.

In addition, we also identify the significant impacts of socioeconomics. Spatial units with relatively long commute times are less likely to have more Uber driver partners, since those commuters are less likely to driver by themselves in Manhattan. If one spatial unit with more workers without available vehicles, it is less likely to provide more Uber services thus fewer Uber supply in outer Manhattan. A contrast case in Manhattan regardless of peak and off peak hours, likely since vehicles are attracted by high service demand and move to those area of interest. In outer Manhattan, we can also capture the negative impacts of residents with Bachelor's degree or higher on Uber supply. One resident with bachelor degree or higher is less likely to serve Uber. However, the employment ratio is different and has significantly positive impacts on Uber supply. High employment rate generally lead to more population movements thus likely to have more available vehicle space for Uber passengers. One unit increase in employment ratio results in more than 0.5 and 1.1 Uber supply during off peak and peak hours, respectively.

In final, the weather of raining will decrease Uber supply by 0.17 during off peak hours, at airports. Moreover, Uber driver partners prefer to online in spatial units where also have high shared bike usage during peak hours in outer Manhattan. They are likely attracted by potential demand of short trips.

4.7 Conclusions

In this chapter, we employs multiple modeling structures for taxi demand and supply specifications, including Poisson regression, simultaneous equation system, spatial lag Poisson, and geographically weighted Poisson regression. The modeling structures are expected to address spatial autocorrelation and modal correlations. Multiple cases are developed with Uber and yellow taxi activities in NYC in April, 2017. In addition to the parameter estimates, we also empirically investigate the appropriate model settings and selections. Based on model performance, we conclude that spatial lag Poisson model performs well enough, compared to geographically weighted Poisson regression. The simultaneous equation system fits yellow taxi and Uber demand worse than independent analyses, although it is designed for correlated dependent variables. Moreover, the model estimation strategy supports the independent analyses for yellow taxi and Uber demand in outer Manhattan, but an overall demand estimation may be better in Manhattan and at airports.

The results identify significant impacts of spatial lag, socioeconomic, land use, transportation built environment, weather, shared bike usage, and dynamic pricing on both demand and supply. The positive spatial lag impacts indicates the taxi activities always concentrate in certain spatial units, which yields spatial units are surrounded by ones with similar activities. Dynamic pricing is effective in controlling demand generation but not effective in attracting new supply. Meanwhile, this study goes beyond the identification of significant variables and introduce marginal effects to validate spatial, temporal, and even modal heterogeneity in impacts.

The proposed modeling structures, as well as model specifications and model outputs, are expected to benefit in multiple ways. First, the modeling structure can estimate not only traditional taxi rides but also the app-based taxi ones. This can be developed as a demand or supply generation tool and included into a taxi planning or urban transportation planning framework. Second, the identified impacts can track the fluctuations of taxi demand and supply, given the changes in the urban social system, which is useful for urban planning. Third, the results also reveal some directions while making policy and improving transportation management. For example, the TNCs should propose strategies to more effective dynamic pricing scheme that is attractive for Uber driver partners. In future study, the work can provide more insights into airport regions if we extend study periods with more hours.

5. AGGREGATED ANALYSES FOR APP-BASED TAXI DRIVERS' PLATFORM-EXITING BEHAVIORS

5.1 Introduction

One of apparent characteristics for app-based taxi services (ATS) is the free entry regulation. In other words, one ATS driver partner can begin or stop their service anytime and anywhere if there are no any passengers on his/her vehicle. This is different from the long-standing taxi shift regulations with minimum working hours (generally 8 hours). The free entry regulation will result in frequent entries and exits. Acknowledging these dynamics is useful for both policymakers and ATS platforms. For example, ATS platforms should know whether their real-time control policies are effective in keeping driver partners active and what factors ATS driver partner value much. Policymakers are also eager for aggregated supply while working on regional transportation planning and policy development. However, the free entry makes this problem more difficult than before, due to involves of numerous individual decisions and large-scale system. Motivated from this point, this chapter focuses on platform-exiting behaviors and devotes to investigate appropriate modeling structures for such behaviors, as well as influencing factors.

Aggregated analyses for platform-exiting behaviors primarily measures portions of empty vehicles exiting platform in a specific region and time. As individual analyses, it also requires high-resolution mobility dataset to track, not only Uber driver partner's exit status but also how many empty vehicles are remaining in the specific region and time. Moreover, ATS driver partners are relatively more informative for surrounding demand and price changes with smart-phone based apps thus are able to make decisions accordingly. This is demanding more endogenous variables to explain their behaviors. On the other hand, the aggregation result in another challenge for complex spatiotemporal autocorrelation patterns, not presented in individual analyses. In general, the both ATS and TTS taxi activities concentrate within city central areas rather than remote ones. The ATS driver partners are also more likely to exit within specific regions and hours, rather than others. Without any addresses for such data structure will result in bias and misunderstands of behaviors.

In this study, we mainly address aforementioned problems, not presented in the current literature, including: 1) are the platform-exiting probabilities influenced by regional app-based platform operation condition? and 2) how the platform-exiting behaviors varies across spatial and temporal scales? We introduce several econometric models to specify the platform-exiting probabilities, including basic modeling of logistic regression and panel analyses, as well as spatial lag model for potential spatial autocorrelations. One month Uber trajectory data is aggregated to quantify platform-exiting probabilities and measure other endogenous variables in Uber system, for instance, demand, supply, and pricing. Except for significant variables influencing platform-exiting behaviors, the modeling structure, as well as their variants, are compared for a better specification.

The remaining sections are organized as follows: section 5.2 introduces data preparation and study case; section 5.3 presents multiple modeling structures for dependent variables of odd ratio, including logistic panel regression and spatial lag adaption; section 5.4 shows the performance of various modeling structures and selects the appropriate one; section 5.5 analyzes the empirical findings from the study case; and section 5.6 concludes our findings and future studies.

5.2 Data Preparation

The platform-exiting probabilities is primarily measured from Uber trajectory data, based on spatiotemporal aggregation scales presented in Chapter 3. First, every trajectory point is labeled with community district (spatial scale), minute (temporal scale), and vehicle id (individual scale). Then, we will count the number of platform-exiting vehicles, as well as empty vehicles circulating under same spatial and temporal scale. Same as demand generation study in chapter 4, this study is also based on two typical hours (off peak from 10am to 11am and peak from 6pm to 7pm) in 9 weekdays of April, 2017 and obtains 12,960 (i.e. 12 Community Districts [CDs] $\times$ 9 days $\times$ 120 minutes), 50,760 (i.e. 47 CDs $\times$ 9 days $\times$ 120 minutes), and 2,160 (i.e. 2 CDs $\times$ 9 days $\times$ 120 minutes) for Manhattan, outer Manhattan, and airport, respectively.

Multiple endogenous variables are also measured from Uber mobility dataset, including number of Uber pickups, average dynamic pricing multiplier, and number of Uber newly online vehicles, under same spatial and temporal scale as platform-exiting probabilities. In addition, we develop an indicator variable of peak hours to indicate whether the minute is within peak hours from 6pm to 7pm and distinguish potential temporal heterogeneity. The summary statistics are presented in Table 5.1. Again, this study has separate analyses for Manhattan, outer Manhattan, airports. Thus, the table has three separate statistics. The multi-collinearity test is employed to filter out independent variables, which is based on variance inflation factor (VIF). In general, the rule of thumb indicates that keeping only explanatory variables with VIFs less than 10 can reduce the dependency. The corresponding results are also presented in this table.

5.3 Methodology

5.3.1 Odd ratios and Logistic Regression

Alternative to direct use of platform-exiting probabilities p_{it}, we prefer to the odds ratio $p_{it}/(1-p_{it})$ and its log-transformation, as equation 5.1. This transformation is the basis of logistic regression model that is widely used for choice problems. The advantages over probabilities are in two-fold: first, the odds ratio represents the constant effect of one explanatory variable x on the likelihood that one outcome will occur. However, there is no way to express such effects with probability; and second, the log-transformation further extends the scale from positive to infinity thus reduces

Table 5.1.: Summary statistics for platform-exiting probability and explanatory variables

Variables	**Unit**	**Min**	**Max**	**Mean**	**Variation**	**VIF**
Manhattan						
platform-exiting probability	n/a	0	0.40	0.009	0.023	n/a
Number of Uber pickups	per minute and CD	0	65	9.37	8.82	1.34
Average surge pricing multiplier	n/a	1	1.8	1.025	0.098	1.22
Number of newly online vehicles	per minute and CD	0	9	0.80	1.023	1.18
Number of empty vehicles	per minute and CD	0	83	27.22	17.87	1.32
Outer Manhattan						
platform-exiting probability	n/a	0	1.00	0.009	0.043	n/a
Number of Uber pickups	per minute and CD	0	18	1.82	1.90	1.15
Average surge pricing multiplier	n/a	1	2.0	1.003	0.034	1.01
Number of newly online vehicles	per minute and CD	0	7	0.35	0.641	1.16
Number of empty vehicles	per minute and CD	0	71	12.64	10.26	1.31
Airports						
platform-exiting probability	n/a	0	0.33	0.009	0.027	n/a
Number of Uber pickups	per minute and CD	0	19	4.54	2.873	1.22
Average surge pricing multiplier	n/a	1	1.6	1.004	0.044	1.02
Number of newly online vehicles	per minute and CD	0	6	0.71	0.922	1.13
Number of empty vehicles	per minute and CD	0	52	18.29	11.78	1.33

model difficulties. Except for the revised dependent variable, we also employ logistic regression (equation 5.2) as a basis one and measure model performance of other proposed approaches.

$$y_{it} = log\left(\frac{p_{it}}{1 - p_{it}}\right) \tag{5.1}$$

$$y_{it} = b + AX_{it} \tag{5.2}$$

where, i represents community district; t represents minute; p_{it} represents platform-exiting probability or ratio of number exiting vehicles over empty ones; y_{it} represents log-transformation of odds ratio, also transformed dependent variables; b represents intercept or constant; A represents a vector of n parameter estimates; and X represents a vector of n explanatory variables at community district i and minute t.

5.3.2 Logistic Panel Analyses and Spatial Lag

However, the standard logistic regression model is inferior to address complicated data structures, such as panel data with cross-sectional and serial observations, as well as spatial autocorrelations. Thus, we further adapt panel data analysis method for such problem. Depending on modeling structures, it has different basic assumptions. Two typical panel analysis methods are fixed and random effects. The fixed one assumes that the unique attributes of individuals (i.e. community districts in this study) do not vary across time and should not be correlated with other individual characteristics. Individual heterogeneity may bias the outcome and should be addressed by way of removing effects of those time-invariant characteristics. Moreover, the individuals' error term and constant variable should not be correlated with others, as equation 5.3. In contrast, the random one assumes that there are unique and time constant attributes of individuals that are not correlated with the individual regressors. In other words, the variation across individuals is assumed to be random and uncorrelated with the explanatory variables, as equation 5.4.

$$y_{it} = AX_{it} + \alpha_i + \mu_{it} \tag{5.3}$$

$$y_{it} = AX_{it} + \alpha + \mu_{it} + \varepsilon_{it} \tag{5.4}$$

where, α_i is the estimated constant for each individual, m in total if there are m individuals; μ_{it} is the between-individual error term; and ε_{it} is the within-individual error term.

Although the panel data analysis model can process complex two-index (one individual and another time) data structure, it is still possible to have cross-sectional dependence, which can be examined with Breusch-Pagan(BP) Lagrange Multiplier(LM) test. Here, we can still refer to spatial lag for corrections, presented in demand generation model of chapter . We skip detail mathematical forms and procedures and just show corrections for panel analysis, as equation 5.5

$$y_{it} = AX_{it} + \alpha_i + \mu_{it} + WY_t \tag{5.5}$$

where, W is distance based weight matrix, as equation 4.4; and Y_t is the matrix of explanatory variables measured at minute t with dimension of $m \times n$.

5.4 Model Performance and Selection

With the study case of Uber in NYC, we estimate proposed models and compare their performance for an appropriate one. Initially, we examine the fixed effect Panel model against logistic regression, as shown in Table 5.2. The OLS part shows the performance of standard logistic regression model and reveals that separate analyses perform better in Manhattan borough and airports. The additional χ^2 distribution based likelihood ratio test provides statistical evidence that separate analyses are better in explaining platform-exiting probabilities. Then we introduce a factor variable of location and check the spatially heterogeneous impacts from the community districts themselves. With this additional variable, almost all regions except for airports have better fitting, which is also confirmed by Anova tests. Additionally, we employ likelihood ratio test method to compare the performance of separate analyses against overall OLS model with spatial effects. The test yields a χ^2 statistic 1280.8 with degree of freedom of 8, thus rejects the null hypothesis that there is no difference between two models. Even with spatial factors, one overall model for all CDs is inferior to fit this problem. This also indicates spatial heterogeneity in large-scale system, in particular between city central and remote areas. In this table, we also estimate a simple one-way fixed effect panel model and utilize F-test to identify its advantages in areas except for airports.

Then, we deeply investigate various modeling structures of panel data analysis methods, as shown in table 5.3. It contains multiple tests, including Hausman test

Table 5.2.: Model performance between panel analyses and ordinary least squares (OLS)

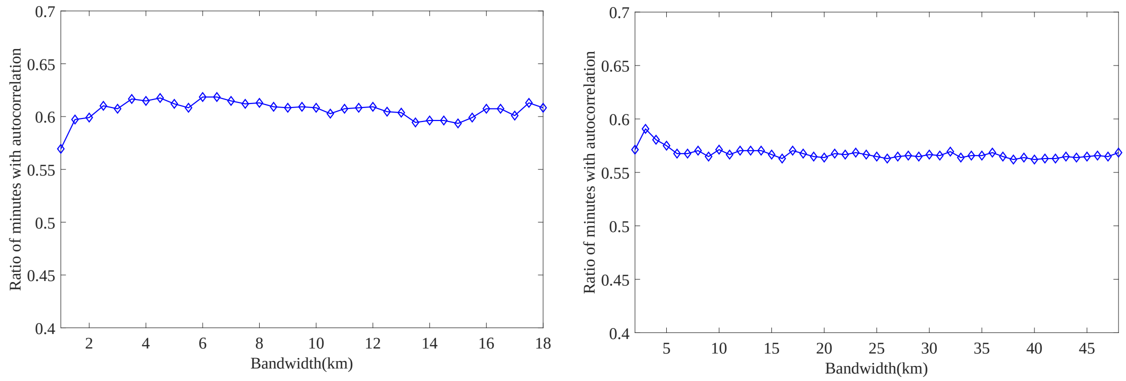
Performance	All CDs	Manhattan	Outer Manhattan	Airports
OLS				
R^2	0.144	0.272	0.092	0.201
Adjusted R^2	0.144	0.271	0.092	0.200
log-likelihood	-153009	-31837	-114648	-5135.6
df	13	11	11	7
OLS-separate vs full				
χ^2 (df)	2776.8 (16)			
OLS with spatial effects				
R^2	0.163	0.277	0.104	0.201
Adjusted R^2	0.163	0.276	0.103	0.200
log-likelihood	-152261	-31794	-114317	-5135
df	66	17	52	7
OLS with spatial effects vs. OLS				
Anova test F-stat (p value)	28.53 (0)	14.27 (0)	16.28 (0)	n/a
OLS separate vs with spatial effects				
χ^2 (df)	1280.8 (8)			
Panel fixed effects vs OLS				
F-stat (p value)	2.00 (n/a)	16.86 (0)	15.69(0)	7.03 (n/a)

for fixed vs. random effects, F-test for one-way vs. two-way fixed effects, and BP LM test for cross sectional correlations. The Hausman test indicates that random effect panel data analysis is not helpful to improve model performance for this study. Within fixed effect panel analysis, we further explore whether we should introduce

Table 5.3.: Model performance between spatial lag and panel analyses

Performance	All spatial units	Manhattan	Outer Manhattan	Airports
Panel -one way fixed effects				
R^2	0.13	0.26	0.08	0.17
Adjusted R^2	0.13	0.26	0.08	0.16
Panel -Random vs. Fixed effects				
Hausman test χ^2 (df),p value	430.49(4), 0	89.08(3), 0	170(3), 0	49.58(3), 0
Panel -Fixed time effects				
R^2	0.17	0.37	0.12	0.59
Adjusted R^2	0.15	0.31	0.10	0.17
F test F-stat,p value	3.05, 0	1.92, 0	2.08, 0	1.02,0.34
Panel -cross sectional correlations				
BP LM test χ^2(df),p value	9557(1830), 0	880(66), 0	4215(1081), 0	0.07 (1),0.79
Panel - spatial lag				
R^2	n/a	0.47	0.74	n/a
Adjusted R^2	n/a	0.47	0.74	n/a
F test with fixed time effects; F-stat,p value	n/a	0.10,1	0.07,1	n/a

fixed time effects and find strong supports of fixed time effects except for airports. The goodness of fit R^2 and adjusted R^2 are also 0.1 more than standard fixed effect panel analysis. In addition, BP LM test reveals potential spatial autocorrelations among community districts. The global Moran's I test on model residuals also provides same evidence in Figure 5.1. There are high portions of periods, over than 55%,



(a) Manhattan (b) Outer Manhattan

Fig. 5.1.: Results of Spatial Autocorrelation Tests in each minute

exhibiting significant spatial autocorrelation, which is much higher than those tests for demand and supply under same spatial and temporal scales. In particular at bandwidth of 6km in Manhattan and 3km in outer Manhattan, the portions reach peak of more than 60%. For such condition, we should introduce spatial lag to correct dependent variables. From Table 5.3, we can see that applying spatial lag can greatly improve fixed effect panel analysis regardless of one-way and two-way, given the comparisons on R^2 and adjusted R^2. The F-test for mixture of spatial lag with two-way panel analysis does not yield a better specification in both Manhattan and outer Manhattan boroughs. Furthermore, we compare geographically weighted panel model with spatial lag panel one. However, same as in demand generation model of previous chapter, it does not lead to better specifications.

To sum up, we can derive following model recommendations:

- Compared to an overall model for all spatial units in a large-scale system, multiple separate models for subdivisions (for instance, city central vs. others) may be better. This is recommended in particular for systems with severe unbalanced activities;

- A two-way fixed effect panel analysis is better than standard logistic regression. The advantage also exists while against random effect and one-way fixed effect panel analysis;
- The log of odds ratio for platform-exiting has significant spatial autocorrelations. A special correction should be applied for much better specifications.

5.5 Empirical Results and Discussions

After extensive comparisons on various modeling structures, we estimate three separate models for Manhattan, outer Manhattan, and airports, respectively. The mixture of spatial lag with one-way fixed effect panel analysis is proposed for Manhattan and outer Manhattan. Moreover, a fixed effect panel analysis model is proposed for airports.

Table 5.4 summarizes all significant impacts of endogenous variables. However, there are almost no effects generated by endogenous variables in all community districts except for airports. In Manhattan and outer Manhattan, only spatial and temporal factors can account for more than 50% variations. The significant spatial lag indicates that Uber driver partners tend to exit in similar community districts. Once one 'weighted' exit in surrounding community districts, there is also 197% and 189% more odds ratio of having additional exits in the community district of Manhattan and outer Manhattan, respectively. Moreover, we can observe positively time-of-day impacts on platform-exiting behaviors. During peak hours, through generally with high demand, Uber driver partners also tend to exit and the odds ratio of having more exiting increase by at least 8%. Surprisingly, the Manhattan borough has a relatively higher increase in odds ratio during peak hours. This is likely due to high portions of part-time drivers, who just serves several trips then exits or pickup passengers en-route to personal destinations.

The airports present significant impacts of Uber pickups, surge pricing, and newly online vehicles. However, it seems that Uber control policies are not effective in

Table 5.4.: Estimation results for Uber exit probabilities

variables	Para.	t-stat	Odds Ratio
Manhattan			
spatial lag	1.09	71.87	2.97
peak hours	0.38	5.31	1.46
outer Manhattan			
spatial lag	1.06	358.83	2.89
peak hours	0.08	6.71	1.08
Airports			
Number of Uber pickups	0.09	4.01	1.09
Surge Pricing Multiplier	3.50	2.65	33.12
Number of Uber onlines	0.17	2.66	1.19
Peak hours	-2.31	-20.02	0.10

keeping Uber drivers active. On the other hand, Uber driver partners are not chasing high incentives from Uber and do not make platform-exiting decisions solely based on earnings. More pickups and higher surge pricing multiplier can generally lead to more revenue, but also increase odds ratio of exiting platform. The newly online vehicles have positive impacts on exiting with an increased odds ratio by 19%, likely due to increasing competitions among driver partners. Last, peak hours can keep driver partners active at airports, given the negative impacts and 90% reduction in odds ratio.

Figure 5.2 primarily shows the fixed location effects from panel analysis. The results reveal a contrast pattern against demand distribution. As one community district has more demand or taxi activities, it is less likely to keep driver partners active. Conversely, remote community districts have relatively smaller increases in

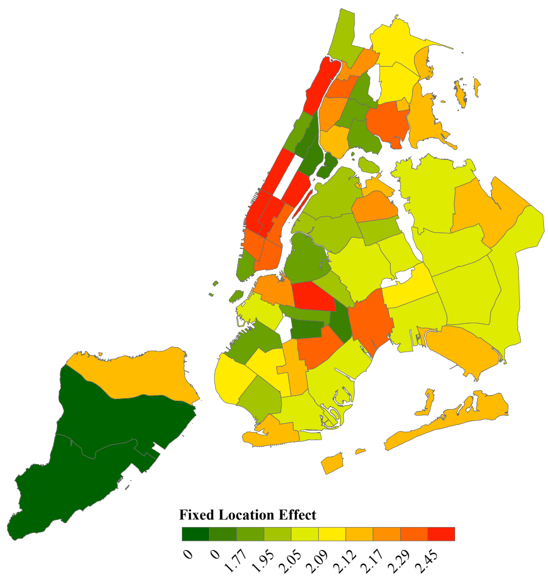


Fig. 5.2.: Fixed location effects on platform-exiting probabilities

odds ratio of having more exiting. Moreover, there are very few community districts having negative impacts and reductions in odds ratio of exiting, but primarily locating to remote areas of each borough.

In the taxi system modeling, the drivers' platform-exiting behaviors are crucial components, determining available supply. Our empirical findings confirm the existence of spatial heterogeneity and autocorrelation in aggregated platform-exiting behaviors. However, our analyses can not provide strong evidence of drivers' rationality and effectiveness of dynamic pricing on supply increases. With high-resolution mobility dataset, we quantify the performance of various modeling structures and their efficiency in addressing complicated spatiotemporal relationship. Another implication of the study sheds light on the predictions of taxicab supply. The proposed methods can simulate drivers' platform-exiting behaviors, quantify hazards of exiting, and predict supply for local regions. This is critical for local transportation system and facility planning.

5.6 Conclusions

The objective of this study is to empirically discover how Uber driver partners stop working and exit at aggregated spatial and temporal levels, as well as identify what regional attributes influence the aggregated behaviors. All analyses are derived from fixed effect panel analysis, and spatial lag corrections for autocorrelation at some regions. Moreover, we calibrate adaptive bandwidth for spatial lag corrections. In the both study cases, we also quantify model performance and propose modeling structure recommendations among various panel analysis, logistic regression, and spatial lag. Except for spatial heterogeneity and autocorrelation in this study, one interesting but also surprising finding is that Uber driver partners are not chasing high surge pricing multiplier, as well as high demand and peak hours.

6. EXPLORING THE DYNAMICS OF SURGE PRICING IN APP-BASED TAXI SERVICES

6.1 Introduction

Recent years have witnessed a revolution in the ATS with the emergence of a large number of TNCs such as Uber, Lyft, Didi, and Ola. Compared to TTS, ATS is efficient in optimal driver-passenger matching and in contrast to the fixed fare rates of the TTS, provide a flexible fare structure to offset the demand-supply mismatch. During high demand at specific location and time, ATS charges higher prices by multiplying the standard fares in normal hours by a *surge price multiplier* (SPM). As claimed, higher SPM encourages more drivers to join the market, thus increasing supply as well as decreases demand since people who can wait for a ride tend to wait till the prices return to normal. This balances the demand and supply and decreases waiting time. While ATS is becoming extremely popular, concerns over their fairness, efficiency, social responsibility, and economic impacts are rising. One particular aspect for which ATS is usually criticized is very high SPM during large public events and emergencies. For instance, Uber was condemned for high SPM (as high as 7) during Hurricane Sandy in New York City (NYC) [152], during a heavy snowstorm which coincided with the New Year's Eve 2011 in NYC [153], during hostage situation in Sydney, Australia in 2014 [154], and most recently during London terror attack in June 2017 [155]. Although Uber tried to do some damage control by reaching a deal with NYC on surge pricing in emergencies [156], which was later extended to the entire US [157], the measures were mostly ad-hoc and not universal.

Opacity of the SPM algorithm and the lack of extensive data driven studies to back the hypothesis that SPM "acts as a tool to balance demand supply mismatch" have made ATS prone to criticism on the rationality of dynamic pricing. Understanding the underlying mechanism behind the dynamic pricing will not only provide the rationality for this scheme but will also help model urban mobility to yield sustainable regulation policies for a unified taxi market consisting of both TTS and ATS. Two major hurdles pose difficulty in providing hard evidence in favor of rationality of dynamic pricing: 1) lack of quality data availability to independent researchers and 2) complex spatiotemporal patterns of the dynamic pricing at a city-level study area, which are challenging to model.

In this chapter, we mine spatially and temporally rich driver trajectory and SPM dataset to empirically explore the dynamics of SPM. Mining this dataset enables us to observe when, where, and how much the SPM is for different products across different locations in the NYC, determine whether ATS is competitive compared to TTS even under SPM, and uncover the underlying mechanism of SPM generation. The main contributions of this study are as follows:

- We analyze the dynamics of SPM on a city scale which enables us to identify the complex spatiotemporal distributions of dynamic pricing.
- We explore the underlying SPM generation mechanism and figure out what contributes to SPM among demand, supply, day and time, and ETA.

For the rest of the paper, we analyze the spatiotemporal distribution of SPM and its effectiveness in section 6.2. Section 6.3 uncovers the underlying mechanism of SPM generation, followed by the summary of the findings in section 6.4.

6.2 Common SPM Patterns

Dynamic pricing implemented in the form of SPM is a major distinguishing factor between ATS and TTS. While the fare of TTS is fixed based on the distance traveled and the travel duration (excluding fees for special locations such as airport and tips), the price for a Uber ride is based on variables subject to change over time [158]. Apart from the estimated time and distance of the predicted route, the fare for a

Uber ride depends on a variety of other factors such as estimated traffic, and the number of riders and drivers using Uber at a given moment. When demand for rides exceeds driver supply, Dynamic pricing help restores the demand-supply balance by temporarily increasing the fares. Surge pricing has two effects: people who can wait for a ride often decide to wait until the price falls and it encourages more drivers to get on the road and head to those areas of the city where SPM is high.

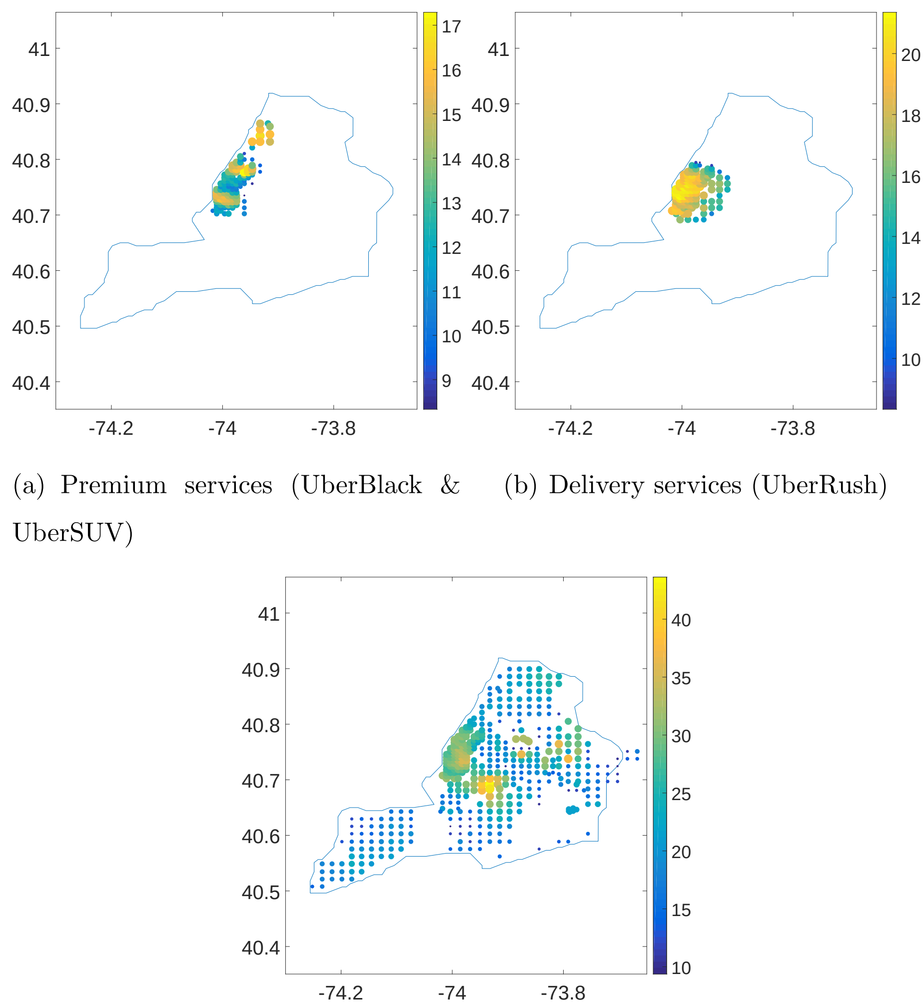
Our analysis of the SPM and ETA dataset found that SPM for different services belonging to same category (e.g. premium, economy, and delivery) are same, even though the number of drivers and potential riders for those services could be different. Based on the similarity of the SPM over location and time, the Uber services in NYC can be categorized into following three categories:

- Premium services: UberBlack & UberSUV
- Economy services: UberX, UberXL, UberFamily, and UberWav
- Delivery services: UberRush

This observation is similar to the observation made in [159] (although for UberX and UberXL only). Many drivers are logged-on on multiple services of the same category, hence the supply of drivers for different products of the same category is quite similar and could be a possible reason for the same SPM. Another logical reason as explained in [159] is that "Uber does this so that riders do not take advantage of the different surge prices across different product lines at the same point of time." For the following sections, we only analyze three representative products for each category, i.e., UberBlack for premium services, UberX for economy services, and UberRush for delivery services.

6.2.1 SPM: where, when, and how much?

Next, we find the common patterns in the SPM by location and time. For each of the three services, we divide the SPM and ETA dataset by the hour of the day and the day of the week for each location. We then find those time intervals (a given hour of a given day) for which $SPM > 1$ is applicable to at least 10% of the observations in that time interval. Figure 6.1 shows the locations (color coded by the percent of the time), where $SPM > 1$ is applicable for the three services: premium, delivery, and economy. Figure 6.2 shows the average value of the SPM by the hour of the day and the day of the week.

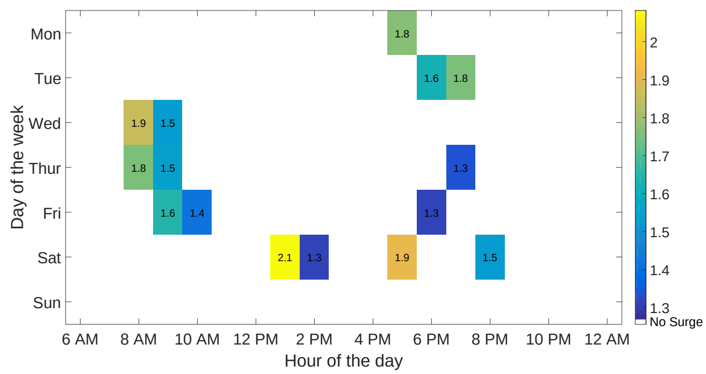


(a) Premium services (UberBlack & UberSUV)

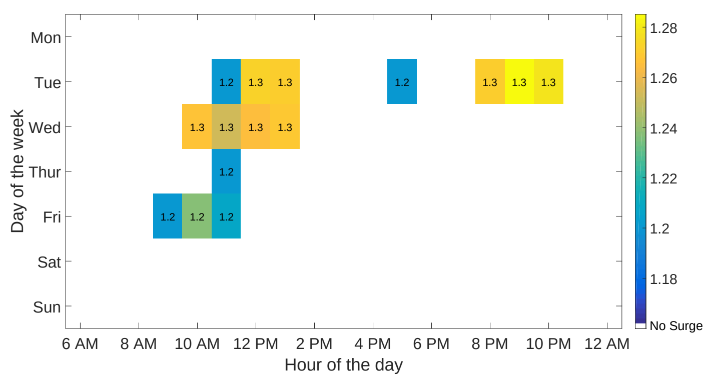
(b) Delivery services (UberRush)

(c) Economy services (UberX, UberXL, UberFamily, & UberWav)

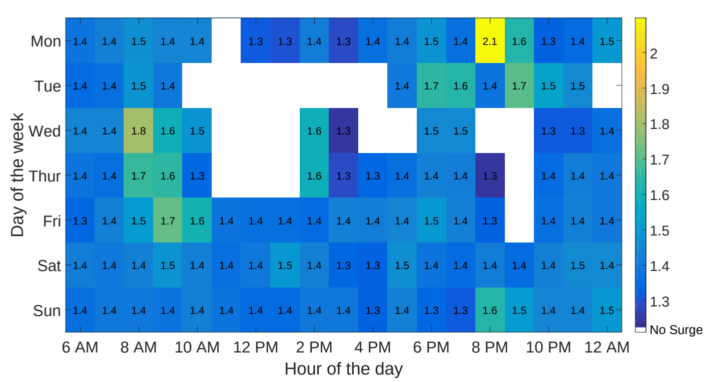
Fig. 6.1.: Percentage of minutes with surge pricing in one hour



(a) Premium services (UberBlack & UberSUV)



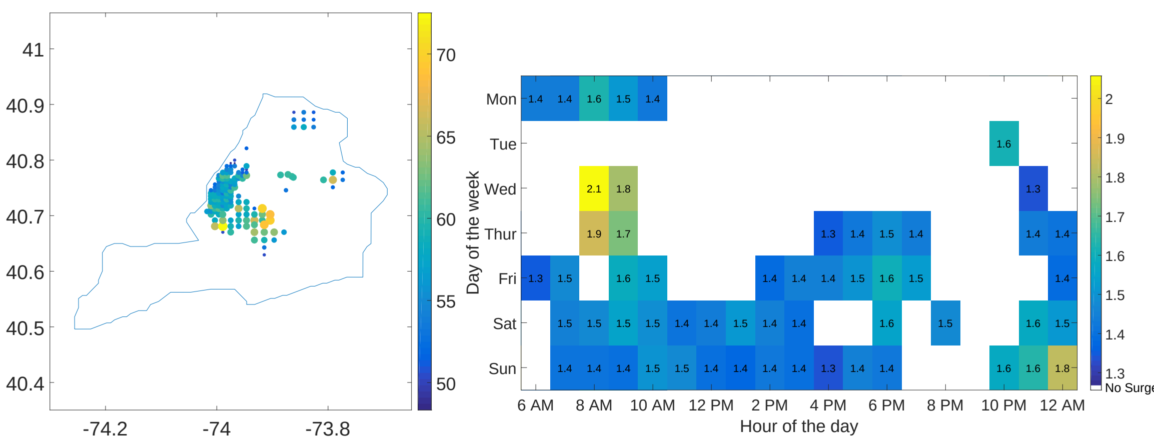
(b) Delivery services (UberRush)



(c) Economy services (UberX, UberXL, UberFamily, & UberWav)

Fig. 6.2.: Average spm based on time of day and day of week

Premium services: UberBlack and UberSUV have a high SPM limited to parts of Manhattan and some areas in the Bronx, with some areas have $SPM > 1$ applicable for as much as 17% of the time during a particular hour of a particular day. Figure 6.2(a) shows the SPM values in the range of 1.4 to 1.9 during morning peak hours (8-10 AM) on Wednesdays, Thursdays, and Fridays. Similar high SPM patterns are also visible during evening peak hours (5-8 PM) on almost all the days. Saturdays afternoon (1 PM) also witness a high surge value of 2.1 due to high lunch-hour demand.



(a) Surge time percent

(b) Average surge value

Fig. 6.3.: Locations and average SPM for economy services when surge is applicable to at least 50% of the time during a particular hour of a particular day

The delivery service UberRush (Figure 6.1(b)) has a similar location profile (concentrated in Manhattan) for the SPM as the premium services since business are the common customer base for both. The temporal profile of UberRush SPM, however, differs significantly from the premium services and has a high SPM applicable on Tuesdays and Wednesdays noon time (11 AM - 1 PM), Tuesdays night time (8 - 10 PM), and Friday mornings (9 - 11 AM).

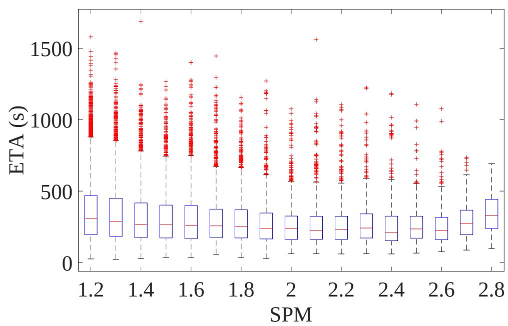
The economy services, UberX, UberXL, UberFamily, and UberWAV, being highly popular, have a much broader spatiotemporal surge footprints (Figure 6.1(c) and

Figure 6.2(c)). Places having surge being imposed for at least 10% of the time for some hours on some days covers the entire NYC area with longer surge durations at Manhattan, the two airports and parts of Brooklyn and Queens. A nominal surge of 1.3-1.4 is applicable to most of the time except afternoon hours (11 AM - 4 PM), Monday through Thursday. Higher SPM is applicable on Weekdays morning and evening peak hours. Since the surge price spatiotemporal patterns, when $SPM > 1$ is applicable to at least 10% of the time during a particular hour of a particular day for economy services covers the entire spatiotemporal domain, we increase the time duration threshold to 50% to find which areas have $SPM > 1$ for more than half of the time during which days and hours. The results are shown in Figure 6.3. These high probability SPM affected areas are mostly concentrated in Manhattan, Brooklyn, La Guardia airport, and some parts of Queens and the Bronx during morning and evening peak, and late night hours on most weekdays and almost entire day during weekends. The highest SPM (~ 2) is applied during morning peak hours (8-9 AM) on Tuesdays and Wednesdays and Sunday late night hours (11 PM - 12 AM).

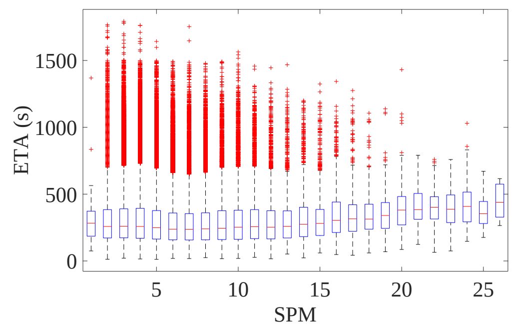
6.2.2 SPM versus ETA

ETA is a comprehensive indicator representing surrounding demand-supply balance. Higher ETA indicates supply shortage since there are no drivers nearby and it would take more time for the nearest available driver to drive up to the pickup location. SPM is designed to minimize ETA by attracting more drivers to the high surge area, meanwhile reducing non-urgent trips. In this analysis, we explore the relationship between the SPM and the corresponding ETA during the surging periods. Figure 6.4 shows the box-plot of ETA for different values of SPM for different product categories, i.e. UberBlack for premium services, UberX for economy services, and UberRush for delivery services.

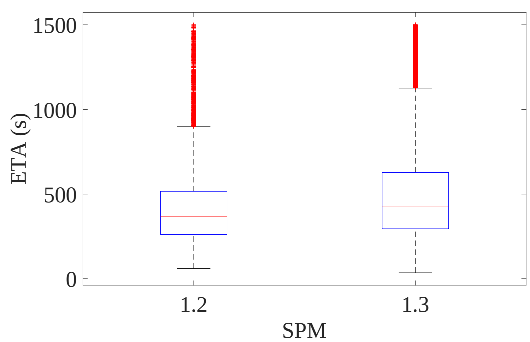
All the three products show significant positive correlations, implying higher ETA leads to higher SPM. However, there exist few differences among products. First, for UberX, SPM increases gradually as ETA increases, but for UberBlack, a very little increase in ETA leads to a higher increase in SPM. Second, UberRUSH being a delivery service, has very small SPM as compared to human mobility services. From Figure 6.4, we can approximate the highest acceptable ETA when no surge pricing is applied: 200 $\sim$ 400s for human mobility services and 400 $\sim$ 500s for delivery service. If an increase in demand or shortage of supply leads to higher ETA, SPM is likely to kick in to restore the demand-supply balance.



(a) UberBlack



(b) UberX



(c) UberRush

Fig. 6.4.: Box-plot of ETA for different values of SPM

6.2.3 How much surge is too much?

The nominal fare (when no surge is in place) for Uber's economy services is generally less than that of the competing TTS and in some cases, UberX is up to 40% cheaper than the local taxicab alternative [160]. However, the application of high SPM could make it more expensive than the street hailing taxis, and sometimes, even the premium service, UberBlack. In this section, we explore the SPM and ETA dataset to find those locations and times, when it is economical to hire a Yellow taxicab or even UberBlack than booking a UberX ride. The base fare (initial fixed charge) and price per mile and minute for UberX, UberBlack, and NYC Yellow taxicabs (all having a capacity of 4 passengers) is given in Table 3.1.

Figure 6.5 shows the fare for UberBlack, UberX, and Yellow taxicab rides for different trip distances (in the range of 0.1 to 20 miles) and trip times (in the range of 0 to 90 minutes). It is apparent that except for very short rides, UberX is economical than Yellow taxicabs and UberBlack is expensive than both of them. During surge times, however, a UberX ride could be expensive than a Yellow taxicab ride, and in some extreme cases, even expensive than the UberBlack ride. We define the *critical SPM ratio* (CSR) for a particular trip distance and time as the ratio of SPM for UberX and UberBlack (for UberX-UberBlack comparison) and UberX SPM (for UberX-Yellow taxicab comparison since for Yellow taxicabs, $SPM = 1$), for which UberX is expensive than UberBlack and Yellow taxicab respectively. Figure 6.6 shows the value of CSR for UberX-UberBlack and UberX-Yellow taxicab for different trip durations and trip distances.

For very short rides, the CSR value for UberX-UberBlack is very high (close to 2.6), but for other rides (trip distance ≥ 2 miles, and trip time ≥ 10 minutes), CSR lies in the range of 2 to 2.1. For UberX-Yellow taxicab fare comparison, for all rides except very short rides (trip distance < 2 miles, and trip time < 10 minutes), a UberX SPM higher than 1.3 could make it expensive than the Yellow taxicabs.

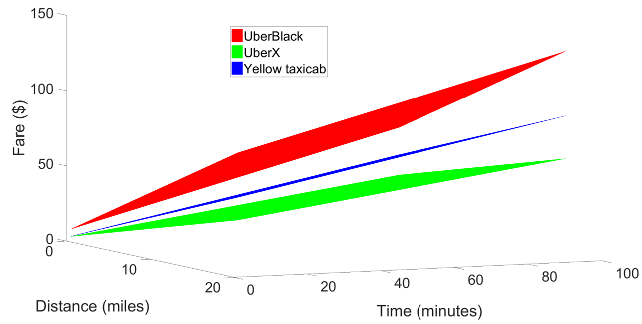
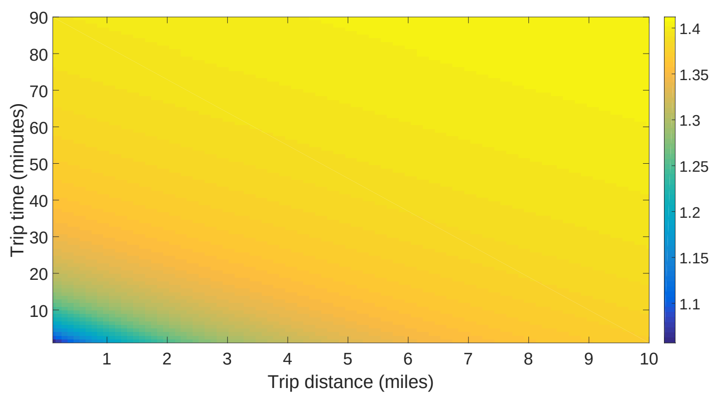
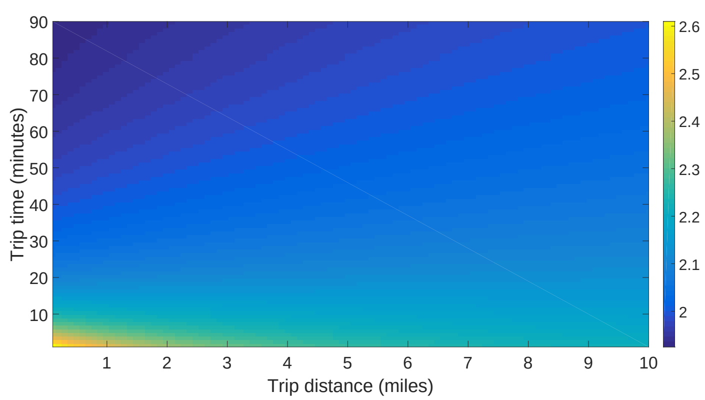


Fig. 6.5.: UberBlack versus UberX versus Yellow taxicab fare for different trip distance and trip time



(a) UberX - Yellow taxicab



(b) UberX - UberBlack

Fig. 6.6.: Critical SPM ratio (CSR): When UberX becomes more expansive than (a) Yellow taxicab, and (b) UberBlack

We use the trips inferred from the gaps in the online driver trajectory data to calculate the average trip distance of 3.2 miles and the average trip duration of 20 minutes. For this trip distance and duration, UberX-Yellow taxicab CSR is 1.35 and the UberX-UberBlack CSR is 2.1. We find those locations and times for which UberX SPM and the ration of UberX and UberBlack SPM is higher than the corresponding CSR values. Tables 6.1 and 6.2 lists the locations and times (the day of the week

Table 6.1.: Fare: UberX > Yellow taxicab

Location	Day	Hour	Probability (%)
La Guardia Airport	Sunday	21-24	43.3
	Mon - Saturday	23-24	15.6
Manhattan-Brooklyn	Wed- Fri	8-9	22
	Thur- Fri	16-19	21.7
	Sat - Sun	13-15	17.7
	Wed - Sat	22-24	17.8
	Monday	6-9	15.6
JFK Airport	Sunday	22-24	19.7
Staten Island	Sat	23-24	17.6
Bronx	Sat-Sun	22-24	14.2
Queens	Sat - Sun	23-24	8.6

Table 6.2.: Fare: UberX > UberBlack

Location	Day	Hour	Probability (%)
JFK Airport	Sun	24	19.6
La Guardia Airport	Sun	22-24	17.3
Manhattan-Brooklyn	Wed- Fri	8-9-10	12

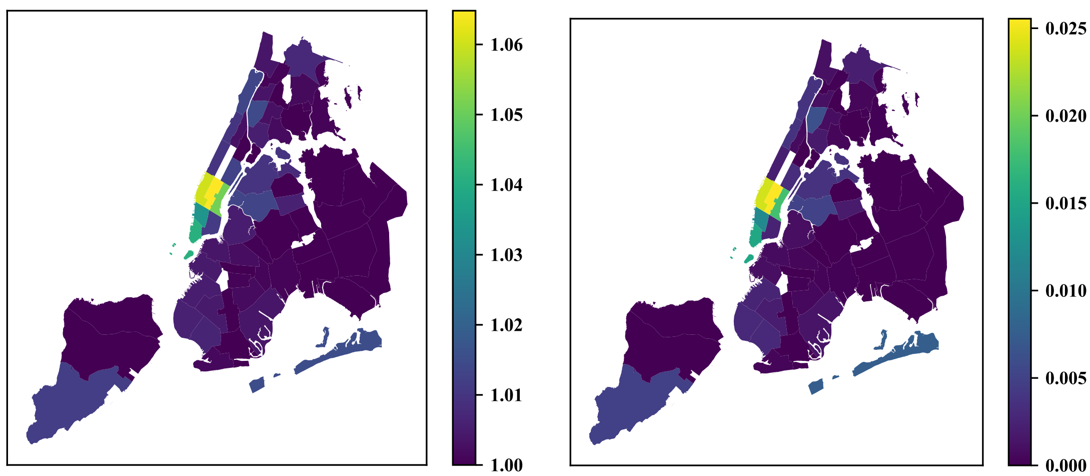
and the hour of the day) for which an average distance and duration UberX ride is expensive than the Yellow taxicab and UberBlack ride respectively. The last column of tables 6.1 and 6.2 shows the probability of this phenomenon happening at the given location at the given day of the week and the given hour of the day. For example, on Sunday nights (9 PM - midnight) at the La Guardia airport, an average UberX ride is likely to cost more than Yellow taxicab ride in 43.3% of the cases. These high surge areas include weekday morning and evening peak and weekend night hours in Manhattan and Brooklyn and weekend night hours at JFK airport, and parts of Staten Island, the Bronx and Queens. It can be inferred from Table 6.2, that during Wednesday - Friday morning peak hours in Manhattan and Brooklyn, and the Sunday night hours at the two airports, the UberX prices surpass that of UberBlack in 12-20% of the cases due to very high surge.

6.3 Dynamic Pricing Generation

6.3.1 Modeling the SPM

We model the SPM in different community districts and minutes for UberX. Since surge rates are charged as a multiplier of 0.1 added to the base surge of 1, this problem belongs to the category of "multi-class classification" in the machine learning domain. The number of classes for the SPM prediction problem is less than 10, varying from 1.0 to 2.0. The average SPM is much less than we experience in reality, since we average SPM in every community district generally much larger than Uber defined hexagon and consisting of multiple hexagons. The mean and variance of SPM across minutes are presented in Figure 6.7. Only city central regions have higher average SPM, as well as variance. All other community districts are generally with SPM of 1.0, as well as very small variations. Thus, it is expect that the dataset is unbalanced and has significant portions of SPM 1.0.

In addition, we aggregate multiple variables in every community districts at the temporal scale of 1 minute. The variables can be grouped into five categories: 1) demand-related variable (D) that is number of Uber pickups; 2) supply-related variables (S) consisting of number of empty vehicles, number of newly online vehicles, and number of offline vehicles; 3) time-of-day related variable (T) that is perk hour indicator (whether the minute is in the hour from 6pm to 7pm); 4) weather related variables (W) that is raining indicator (whether the minute has precipitation); and 5) location relation variables (L) including Manhattan and Queens indicators (whether the community district is in borough of Manhattan or Queens).



(a) Mean surge pricing multiplier across minutes (b) Variance of surge pricing multiplier across minutes

Fig. 6.7.: Statistics over surge pricing multiplier

We use a variety of popular multi-class classification algorithms based on the binary-complete coding scheme, which partitions the classes into all binary combinations and does not ignore any classes. We perform a 10-fold cross validation test to access their accuracy for predicting average SPM. Since most classifiers are highly popular in the machine learning community, we skip detail mathematical formulation and implementation procedures.

Support Vector Machine (SVM)

Given the training dataset of n observations in the form of $(x_1, y_1), \cdots, (x_n, y_n)$, where x_i is a p-dimensional real vector (in our study, 8) and y_i is average SPM in each community districts and minute, the SVM method is designed to find a separate (p-1)-dimensional hyperplane. In multi-class classification problem, the hyperplane fits one class against all other classes, which can be in both linear and non-linear form as equation 6.1 and 6.2. We employ two typical kernels, linear (SVM-L) and radial basis function (SVM-RBF) for SPM problem.

$$f(x) = sign(\sum_{i=1}^{n} y_i \alpha_i < x, x_i > +b) \tag{6.1}$$

$$f(x) = sign(\sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b) \tag{6.2}$$

$$K(x, x_i) = e^{-\frac{|x-x_i|^2}{2\sigma^2}}$$

where, $sign$ is the sign function; α_i is Lagrange multiplier; and $< . >$ is the dot product.

Discriminant Analysis (DA)

The DA method assume that conditional probability density function $p(x|y)$ is normally distributed with mean and covariance (μ, Σ). Then it is built on Bayes' theorem to flip assumed conditional probability around into estimates for probability of y given values of x. Depending on relations of mean and covariance across classes, we have two variants that are linear (LDA) and quadratic (QDA). LDA has common μ and Σ for every class. In contrast, QDA introduce different distribution shape for each unique class.

K-Nearest Neighbor (KNN)

KNN is one supervised machine learning approach. The method is simple, highly accurate, but sensitive to irrelevant features. Moreover, the setting for K may also be problematic. The core of KNN is measurement over data similarity, for instance distance based. Then the algorithm searches over all observations based on similarity and groups similar points into one same categorical class.

Logistic Regression (LOG)

The logistic regression is generally used to predict probabilities of different possible outcomes y, given a set of independent variables x, as shown in equation 6.3. The model can be further generalized for ordinal data (OLOG), such as surge pricing multiplier, through introducing one latent variable (equation 6.4) and multiple thresholds ν (equation 6.5).

$$Pr(y_i = j) = \frac{e^{\beta_j x_i}}{1 + \sum_{j=1}^{J-1} e^{\beta_j X_i}} \tag{6.3}$$

where, j is the categorical class; and J is the number of classes.

$$y^* = \beta x_i + \varepsilon \tag{6.4}$$

$$y = \begin{cases} 0 & \text{if} y^* \leq \nu_1 \\ 1 & \text{if} \mu_1 < y^* \leq \nu_2 \\ \cdots & \cdots \\ J & \text{if} y^* > \nu_J \end{cases} \tag{6.5}$$

Complement Naive Bayes (CNB)

The naive Bayes classifier is a conditional probability model, built upon Bayes's theorem. However, the complement naive Bayes (CNB) adapts for unbalanced dataset (i.e. one or several classes have much more observations against other classes), by estimating feature probabilities for one class with the complement dataset, rather than itself.

$$Pr(y|x) = \frac{p(y)p(x|y)}{p(x)} \tag{6.6}$$

Decision Tree (DT)

The DT method is a flowchart like structure with features as node and outcomes as branches. Thus each leaf node represents a class and internal nodes on that branch form into a classification rule.

We implement all algorithms in Python on an 64-bit Intel i7-6400, 16 GB RAM PC. Multiple performance metrics are measured for model comparisons, primarily based on the confusion matrix of true and predicted categorical classes. The metrics are generalized from binary classification based on one against rest mechanism. True Positive Rate (TPR) measures portions of each categorical class that is rightly labeled. Positive Predictive Value (PPV) measures portions of predicted categorical classes that are real. False Positive Rate (FPR) measures portions of rest categorical classes that are wrongly assigned. Alternatively, we use several additional metrics (see equations 6.7, 6.8, and 6.9), for instance Informedness(INFO), Matthews correlation coefficient (MCC), and G-means (G_M), to obtain comprehensive evaluations at each categorical class. Since F1 score and accuracy are unreliable for unbalanced datasets. In addition, Mean Absolute Percentage Error (MAPE) is computed to check relative errors. Overall, for such unbalanced dataset, it is much more likely to get accurate classification for majority, compared to minority. Depending on MAPE and PPV, CNB yields better classifications than alternative methods, given the facts of smaller relative errors and larger portion of corrected predictive classification. Although QDA performs better at higher SPM, the model is not appropriate for major class of SPM 1.0. Regarding TPR and FPR, we expect better models that are with larger TPR and smaller FPR at all categorical classes. KNN, LOG, OLOG, and SVM-RBF perform better than remains. The comprehensive metrics provide more insights on model performance. OLOG and KNN result in little better results. However, it is hard to compare OLOG and KNN, since KNN works well with smaller SPM but OLOG is appropriate for more categorical classes.

$$INFO = TPR - FPR \tag{6.7}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6.8}$$

where, TP, FP, FN, and TN are confusion matrix elements, indicating true positive, false positive, false negative, and true negative, respectively.

$$G_M = \sqrt{TPR \times (1 - FPR)} \tag{6.9}$$

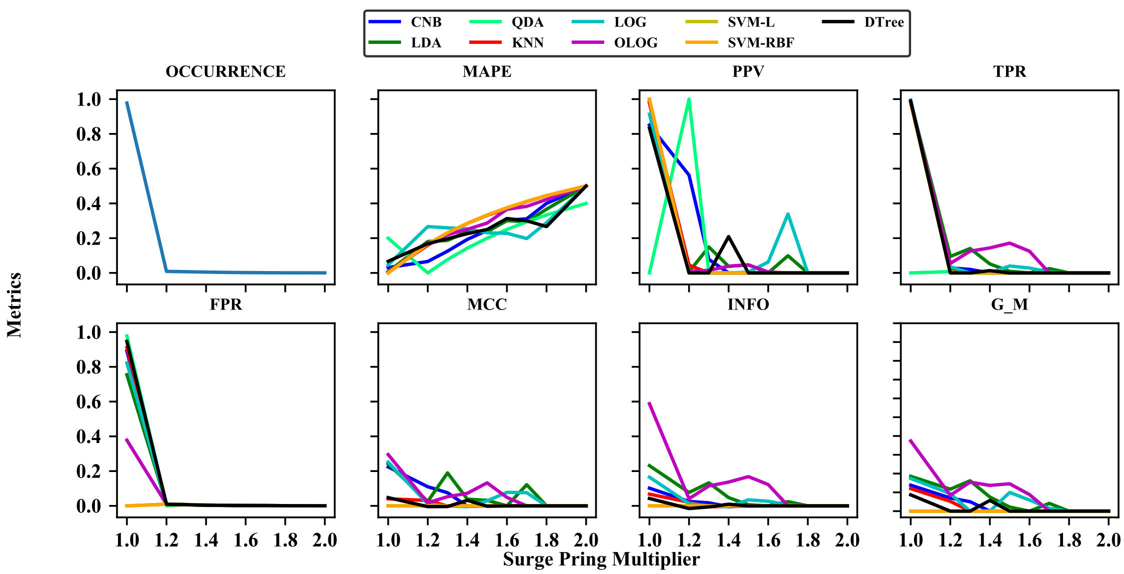


Fig. 6.8.: Model performance with 10-fold cross validation

6.3.2 Critical Factors for Dynamic Pricing Generation

This section continues with one of better classifiers that is ordinal logistic regression. In addition, the method can also provide more parameters of interests, which can help up interpret impacts of variables on probability of employing higher or lower SPM. Figure 6.9 exhibits average model performance (10-fold cross validation) with

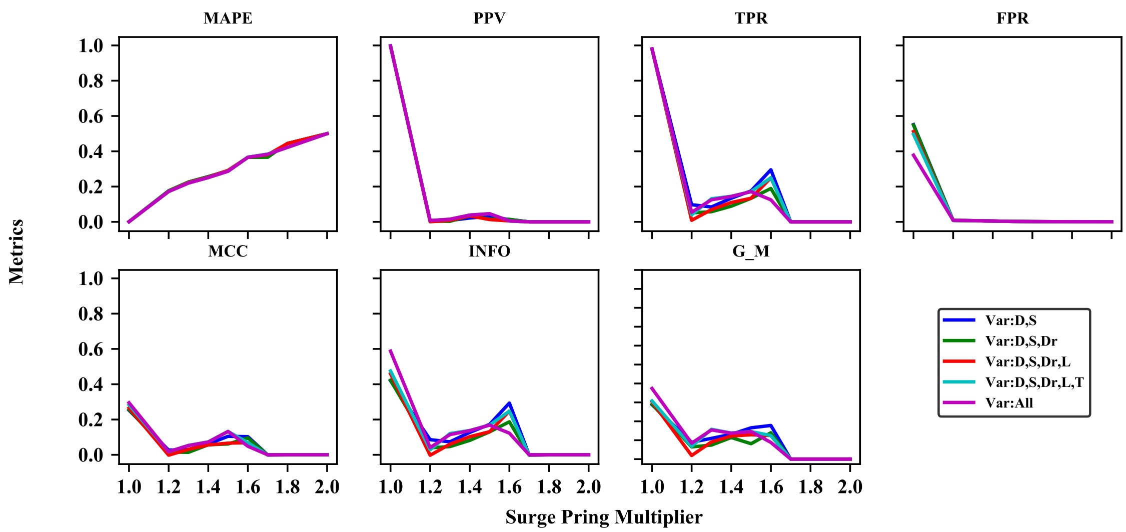


Fig. 6.9.: Model performance with different features (or variables)

different set of variables. The MAPE and PPV do not show significant differences in predictive classification. However, remaining performance metrics reveal several interesting points on variable selections. First, only number of Uber pickups and empty vehicles are enough for higher SPM classification, but are still not good enough to explain how higher SPM issued, considering lower metric values. Second, it seems that the introduction of vehicle exits and online behaviors can not help improve predictive classification by number of Uber pickups and empty vehicles but this can enhance accuracy if combining with additional location, time, and weather features. Third, although the model with all variables can not explain higher SPM well, it reduces the probability of assigning SPM 1.0 to rest categorical classes. Last, only location feature is significantly worse against both spatial and temporal features. To sum up, we can use all variables or features to classify SPM, although they may be a little disadvantage over simple combination of number of Uber pickups and empty vehicles.

Table 6.3 summarize the average estimates from 10-fold cross validation. We can see that except for number of empty vehicles and raining time, all other variables

Table 6.3.: Impacts of variables on SPM from ordered logistic regression

Variables	Group	Para	T-stat	e^{Para}
Uber pickups per minute	D	0.11	31.15	1.12
Number of Newly online drivers per minute	Dr	0.42	12.59	1.53
Number of empty vehicles per minute	S	-0.10	-24.43	0.91
Number of exiting vehicles per minute	Dr	0.47	3.45	1.61
In Manhattan (1 if yes, 0 otherwise)	L	1.70	20.84	5.52
In Queens (1 if yes, 0 otherwise)	L	0.37	3.65	1.53
Raining minute (1 if yes, 0 otherwise)	W	-5.26	-3.83	0.01
Peak hours (1 if from 6pm to 8pm, 0 otherwise)	T	1.76	18.01	6.04

have significantly positive impacts on probability of higher SPM. The last column of Table 6.3 also shows the magnitude of relative risk ratios. In Manhattan, if keeping all other variables unchanged, the odds of posing higher SPM is 452%, against other city boroughs. Similarly, the borough of Queens also has similar higher risk of more Uber payments with an odds of 53%. Another significant variable leading to higher SPM is the peak hours. Against off peak hours, it also has a much higher odds of 504%. On the other hand, the demand and supply have relatively weak impacts. One unit increase in Uber pickups, newly online drivers, and offline drivers lead to 12%, 53%, and 61% odds of having higher SPM in that community district, respectively. In particular for the number of offline drivers, it possibly result in supply shortage thus a higher SPM. Unsurprisingly, number of empty vehicles in one community districts can really reduce SPM. One unit increase in number of empty vehicle can decrease odds by a factor of 0.91. Meanwhile, the raining time also contributes to SPM decreases with a factor of 0.01 reduction in odds. Overall, the SPM is effective in reflecting demand and supply variations, as well as posts much higher fare rates in city central zones.

6.4 Conclusions

In this study, we primarily utilize the data of online driver trajectories and SPM in NYC, crawled from Uber mobile platform. We identified the common patterns in the SPM to find when, where, and by how much is a passenger expected to pay the surged fare. We compared the fare structure and SPM of UberX, Yellow taxicab and UberBlack to find locations and times when UberX fares are higher than that of Yellow taxicabs and even UberBlack due to high SPM in place. We used a variety of multi-class classifiers to model SPM in terms of demand, supply, time of the day, location, and weather. Upon that, we found that kNN and ordered logistic regression are better in modeling the SPM. The findings help understand the dynamics and generative process of SPM and enrich the growing body of literature in this field. This study provides the rationale for the dynamic pricing and could help model urban mobility to yield sustainable regulation policies for a unified taxi market consisting of both TTS and ATS.

7. MODELING URBAN TAXI SERVICES WITH APP-BASED TAXI SERVICES: A QUEUEING NETWORK APPROACH

7.1 Introduction

The rise of TNCs enabled by rapid adoption of smartphones has significantly altered the urban transportation landscape by providing anywhere, anytime mobility. These new mobility services are mainly based on smartphone apps and thus are generally named as e-hailings, app-based taxi service (ATS), or mobility-on-demand (MoD) services. The ATS systems compete with traditional taxi services (TTS) resulting in significant changes in the urban taxi market. For instance, the medallion price of TTS system in NYC has dropped ninety percent in the last five years due to the entry of ATS [161]. The lack of barriers in the new taxi market (a mix of TTS and ATS taxis) leads to an oligopolistic market where a small number of firms dominate the ATS and TTS market. The ATS and TTS vehicles have similar door-to-door mobility services but utilize different hailing methods; and the dominant firms (or taxi authorities) create technical (or monetary) barriers to new entries. The emerging market structure is very different from the previous monopolistic taxi market where all cabs apply for permits from a central taxi authority; have mostly the same service characteristics and pricing schemes and taxi authorities control the fleet size by determining the number of licenses to sell. In particular, there are two fundamental shifts in the ATS market which warrant highlighting:

- Change from a TTS-dominated to a competitive market with both ATS and TTS. In the TTS, the authorities can manage fleet size in a direct and centralized way (e.g. medallions) and has direct interaction with the drivers. On the other hand, the ATS provide a matching platform to pair customers and drivers, and acts as a middleman between riders and drivers;

- Flexibility and Reduced barriers of entry. ATS allows flexible service hours and almost no barriers for entry. One ATS driver can become available for service anytime and anywhere within the city where the vehicle registers. There are no limits on service hours per shift and fleet size unlike TTS. Moreover, the ATS platforms updates the fare rate based on supply and demand conditions instead of fixed fare rate in TTS.

Advances in data science techniques, as well as availability of taxi GPS traces, have contributed to understanding dynamics and operations in taxi markets. In particular, the TTS studies vary from vehicle/passenger movement patterns [100, 162], mode choice [4], service efficiency [5, 22], ride-sharing [145, 146, 163], to system modeling [17, 24, 164]. However, almost all studies focus on one unique taxi service rather than modeling the comprehensive taxi market comprising of both the ATS and TTS, except for a few studies which model the market competitive equilibrium [26, 165]. Given the limited number of studies in the literature on the systematic modeling of the taxi markets, this study will make a contribution to develop fundamental tools to measure system wide performance metrics of the new taxi market.

Currently, there are limited modeling tools that allow for the quantification of the quality of service and capture the inherent stochasticity that arises in the taxi markets. There are three challenges that should be addressed in developing system wide tools for modeling the new taxi markets:

- Spatial Heterogeneity. Since the taxi activities are highly associated with socio-economic variables, the taxi rides also shows significant spatial heterogeneity, particularly, during rush hours. In other words, most of the taxi pickups may be concentrated in some high economic activity zones, such as central business district, while dropoffs are concentrated in other zones. The modeling has to consider properties that dictate the spatial heterogeneity in an urban area.

- Network externalities. Whenever a customer engages a vehicle, this not only decreases the instantaneous vehicle availability at the source location, but also affects the future vehicle availability at all other locations within short timescales. The impacts of network externalities are more significant under dynamic pricing in ATS. Since changes in vehicle availability can be reflected in dynamic pricing this affects future vehicle supply (e.g. induced supply).

- Role of Stochasticity. In a two-sided market, not only do customers choose when to request a ride but also drivers choose when to work, how long to work, and where to search for customers. Moreover, the platforms frequently examine local states and develop corresponding fare rate for a specific time interval and location. This will in turn affect demand and supply. Even if the transportation network is symmetric (i.e. uniform arrival rates and routing choice at all nodes over a regular grid network), the stochastic nature of arrivals will also quickly drive the system out of balance and hence leads to instability.

The large-scale and stochastic nature of the system makes it challenging, but critical and important, to develop an understanding of the system dynamics by considering both ATS and TTS within the same framework. More importantly, the system modeling should be not only rich enough to capture the salient features of both passengers' and vehicles' behaviors but also the stochastic nature of the demand-supply dynamics and the resulting stability of the system. With this background, the goal of this study is to develop a queuing based methodology to model the combined ATS and TTS system (or 'new taxi system') dynamics and determine various performance metrics of the system.

As discussed in Chapter 2, there are four lines of research for taxi system modeling, including aggregated models, equilibrium models, individual based simulations, and queue theoretic approaches. However, most studies are on TTS rather than ATS or both services, except for few adaption of equilibrium models for emerging oligopolistic taxi market with both ATS and TTS. Studies [24, 25] applied the demand-supply equilibrium model into the scenario of the oligopolistic taxi market but assumed the static pricing scheme rather than dynamic one by ATS. Although the study [26] considered dynamic pricing for ATS under a framework of competitive equilibrium between ATS and TTS, they failed to precisely formulate vehicle-passenger matching, as well as utilities of passengers and drivers.

In this study, we take advantage of both queueing network models and matching queues to investigate the large-scale and stochastic nature of the competitive taxi market, yielding quantitative performance measures. Within the queueing network, we have two types of queues/nodes representing two different subsystems. The first type of nodes are the taxi passenger-vehicle matching subsystem. Since neither of the ATS and TTS have stations, we assume that one homogeneous spatial unit is modeled as one taxi subsystem. Vehicles are matched with passengers based on a synchronization process, forming a 'synchronized' $SM/M/1$ queue. The latter differs from the regular $M/M/1$ queue in that it has two independent arrival flows of both passengers and vehicles thus processes synchronized passenger-vehicle pairs that match based on arrival sequences and zero matching time. Although there exist certain differences, the $SM/M/1$ queue can be closely approximated by the simple $M/M/1$ queue by taking the minimum of the two arrival rates as the effective input rate; this approximation is further detailed below.

The second type of nodes are the road transport subsystem. Again the road network is split into homogeneous units and each homogeneous unit is modeled as one road subsystem in the form of an $M/M/c$ queue. The road subsystem can be connected with both road and taxi subsystems based on geographical contiguity. However, the taxi subsystems only serve their spatial unit and can not directly interact with other taxi subsystems, because vehicles departing from taxi subsystems should travel through road network to get to their destinations. Further, we also model balking behavior of vehicles in the $SM/M/1$ queue given the fact that empty vehicles can drive to another spatial unit if s/he can not find passengers at one spatial unit. Finally, we also account for the dynamics of matching efficiency with state-dependent service rates. Note that one of the novelties of our model is that we explicitly combine a model of 'virtual' infrastructure (the passenger-vehicle matching queue) with a model of 'physical' infratructure (the urban road network) to obtain a holistic view of the taxi system dynamics.

Queueing theoretic approaches provide insight into system performance under a range of workload conditions. In particular, it allows us to identify the degree of load that will cause a system to become highly congested without actually cascading into failure. On the other hand, queueing approaches have a reputation for making unrealistic distributional assumptions and of lacking robustness to divergence of the actual system from modeling assumptions. Here, we use multiple statistical hypothesis tests at various spatio-temporal resolutions to justify our queueing theoretic model. In general, queueing theoretic models are data agnostic, and provide sufficient conditions under which one can compute performance metrics of interest. Having said that, it is not *a priori* apparent that a given queueing network is appropriate at a given spatio-temporal resolution, for the taxi system. We combine the queueing network model with extensive statistical hypothesis testing to justify an appropriate spatio-temporal aggregation scale at which the observed arrival and service conditions are sufficiently 'homogeneous,' thereby yielding empirical validation for our modeling assumptions.

The following sections are organized as follows: section 7.2 presents the modeling structures for the oligopolistic taxi market, as well as empirical settings; section 7.3 investigates the stationary state distributions and asymptotic behaviors under the proposed modeling structure; section 7.4 provides a case study based on TTS and Uber in NYC; section 7.5 concludes the study and points out future studies.

7.2 Modeling Structures

7.2.1 Network Presentation of the Competitive Taxi Market

In the TTS-dominated taxi market, the dynamics can be simply explained by a bilateral passenger-vehicle matching, as well as vehicle movements among spatial units.

Symbols	Descriptions
λ_i^p	The total passenger arrival rate of both the ATS and TTS at spatial unit i, and superscript p denotes passengers
p_i^*	The probability of passengers using service * at spatial unit i, and superscript * can be either ATS or TTS
$\lambda_i^{p,*}$	The passenger arrival rate at spatial unit i, superscript p denotes passengers, and superscript * can be either ATS or TTS
$\lambda_i^{v,*}$	The external arrival rate of vehicles at spatial unit i, V denotes vehicles, and superscript * can be either ATS or TTS
$\lambda_i^{pv,*}$	The arrival rate of passenger-vehicle pairs at spatial unit i, superscript pv denotes passenger-vehicle pairs, and superscript * can be either ATS or TTS
$\hat{\lambda}_i^{v,*}$	The effective arrival rate of vehicles at spatial unit i, and superscript * can be either ATS or TTS
μ_i^*	The service rate for passenger-vehicle pairs at spatial unit i, and superscript * can be either ATS or TTS
D_i^*	The departure flow rate of passenger-vehicle pairs at spatial unit i, and superscript * can be either ATS or TTS
λ_i^r	The overall vehicle arrival rate for road queue at spatial unit i, and superscript r denotes the road queue
$p_{i,\cdot}^{r,*}$	The portion of vehicle flows at spatial unit i, subscript . can be either O (occupied vehicles) or E (empty vehicles); and superscript * can be either ATS or TTS
μ_i^r	The service rate at road queue of spatial unit i, and superscript r denotes the road queue

Symbols	Descriptions
$p_i^{p,*}$	The probability of empty vehicles successfully picking up passengers at spatial unit i, superscript * can be either ATS or TTS, and superscript p denotes passenger pickups
$p_{ij}^{\cdot,*}$	The probability of vehicles moving from spatial unit i to j, superscript . can be either O (occupied vehicles) or E (empty vehicles); and superscript * can be either ATS or TTS
$p_{i0}^{\cdot,*}$	The probability of vehicles exiting system at spatial unit i, superscript . can be either O (occupied vehicles) or E (empty vehicles); and superscript * can be either ATS or TTS
$F_{i,in}$	The incoming vehicle flow rate at spatial unit i, regardless of service types and vehicle status
$F_{i,out}$	The outgoing vehicle flow rate at spatial unit i, regardless of service types and vehicle status
I	The set of spatial units with cardinality of $\|I\|$
a_{ij}	The connectivity between spatial units i and j with physical road network
x_i^*	The number of vehicle in taxi queue at spatial unit i, and superscript * can be either ATS or TTS
$x_i^{\cdot,*}$	The number of vehicle in road queue at spatial unit i, superscript . can be either O (occupied vehicles) or E (empty vehicles); and superscript * can be either ATS or TTS
c_i	Number of road servers at spatial unit i

In particular, the passenger-vehicle matching behaviors are critical to the system performance, for instance, waiting/searching time and utilization. Since the existence of spatiotemporal mismatch between one drop-off and the next pickup, taxicab drivers always search around for passengers. Although ATS also operates like TTS, the centralized platform with real-time controls introduces more complexity, as shown in Figure1.3. The first addition is the competition for passengers between ATS and TTS. The second addition is the ATS drivers' flexible working hours. Moreover, the ATS platform examines demand and supply frequently and utilizes dynamic pricing for seeking balance of demand and supply. Overall, the competitive taxi market receives two types of external flows: passengers and vehicles. The vehicles are operated by both the taxi fleet and ATS driver partners. Due to the market model of free entry, the ATS vehicles can enter and exit the system in a frequent manner. Thus, this is an open system involving external arrivals and exits.

As mentioned before, the major behaviors of the taxi system are passenger-vehicle matching and occupied/empty vehicle movements over urban road network. Thus, the analyses should involve not only the taxi system itself but also the urban road network. It is well known that the both systems are spatially unbalanced, for instance downtown with high ride requests, but with slower ground traffic speed. We first divide the whole system (e.g. a city) into multiple subsystems (e.g. homogeneous spatial units). The empty vehicles can meet passengers in each subsystem, called taxi subsystem and then the matched pairs of passengers and vehicles (or occupied vehicles) travel within and across road subsystems. Therefore, it forms a system of systems $G(N, A)$, where N is the combination of all divided subsystems (or spatial units), each of which operates one taxi and one road subsystem; and A is the set of directed links indicating the connections across subsystems, consisting of a_{ij}. Moreover, the directed links are weighted with routing probabilities p_{ij} to describe the routing choices by vehicles. We further classify into four routing probability matrices that all can be derived from our empirical datasets, depending on service types (ATS or TTS) and vehicle status (occupied or empty). Regarding each unique spatial unit, Figure 7.1 illustrate major

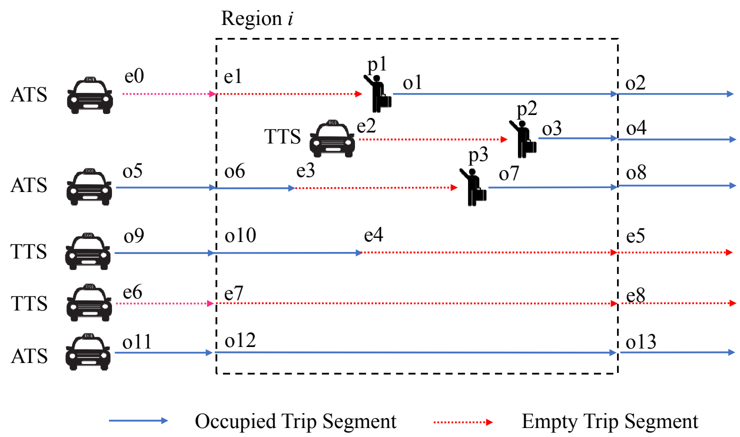


Fig. 7.1.: Illustration of major taxi activities in one spatial unit

taxi activities and segment based on vehicle status. One spatial unit generally receives two external arrivals of both vehicles and passengers (e.g. p1,p2, and p3). In specifics, external vehicle arrivals may originally generate within the spatial unit (e.g. e2) or transfer from neighboring spatial units, regardless of occupied (e.g. o6, o10, o12) and empty (e.g. e1, e2, e7) vehicles. The detail structures of taxi subsystems will be presented in the next section, addressing not only two external arrivals but also more complicated behaviors of dropoff followed by pickup, for instance o6 and e3. On the other hand, each road subsystem only describes the vehicle and taxicab movement over road network directed by the corresponding routing probabilities. In addition, we introduce a virtual node N_0 as the exit node from the system and describe the exiting vehicle flows. In summary, each taxi subsystem only addresses the matching dynamics between empty vehicles and passengers and then transmits matched pairs of passengers and vehicles to the road subsystems. Each road subsystem moves both occupied and empty vehicles among taxi subsystems. In the following two sections, we model each subsystem using queueing models.

7.2.2 Passenger-Vehicle Matching

The taxi system, regardless of ATS or TTS, require matching passengers and vehicles. Standard approaches for modeling matching include nearest distance and Cobb-Douglas production function. However, the former is inappropriate, since it has been observed that even when drivers have perfect knowledge, they do not apply a nearest distance heuristic to find a passenger. On the other hand, it is typically very hard to calibrate Cobb-Douglas production function from available data. A more appropriate approach would be to use matching or assembly-like queues [166]. Here, passengers and vehicles are queued up in separate "buffers" and are matched on a first-come-first-served (FCFS) basis. The arrival flow of passengers and vehicles is determined by a "synchronized" stochastic process, defined as the minimum of the individual flows. Assuming that the individual arrival processes are Poisson processes and that the 'service' (i.e. matching time) times are exponentially distributed, we model the matching process as a $SM/M/1$ assembly-like, synchronized queue. As noted above, we assume that matching is conducted on a FCFS basis, which is reasonable way in which passengers and vehicles are matched. The service time, on the other hand, models how quickly a vehicle can reach a passenger's location within a subsystem. This is critical for modeling the dynamics of ATS, in particular.

Since the ATS and TTS coexist and compete in each subsystem, we introduce the $SM/M/1$ for each of ATS and TTS and deploy a parallel layout besides the interactions while demand splitting, as shown in Figure 7.2. Given a taxi subsystem $i \in I$, the overall passenger arrivals to this spatial unit follow a Poisson process with rate λ_i^P. With Bernoulli splitting, the passengers are split into two Poisson processes with rates $\lambda_i^{P,ATS} = p_i^{ATS}\lambda_i^P$ and $\lambda_i^{P,TTS} = p_i^{TTS}\lambda_i^P$ with $p_i^{ATS} + p_i^{TTS} = 1$. The available ATS vehicle arrival $\hat{\lambda}_i^{V,ATS}$ consists of two sources: (1) the Poisson process of newly joined vehicles (e.g. e2 in Figure 7.1) with a rate $\lambda_i^{V,ATS}$; and (2) empty vehicles who are searching for passengers and successfully pick up in final, but originate from neighboring spatial units (e.g. e1 and e3 in Figure 7.1), $F_{i,in}p_i^{p,ATS}$.

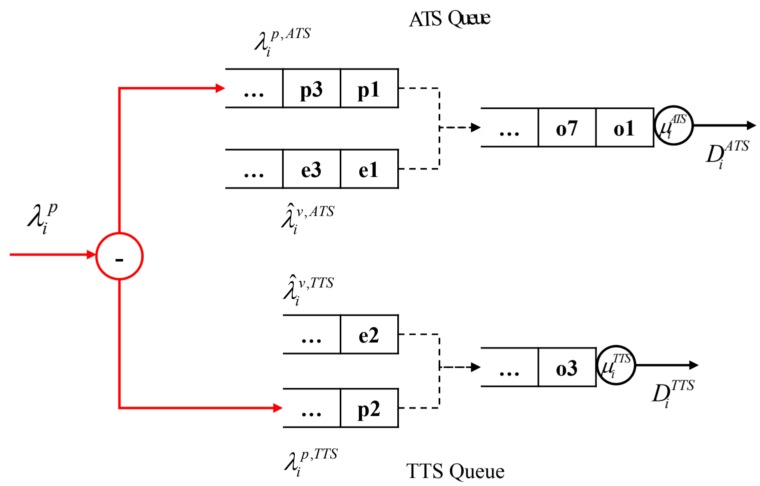


Fig. 7.2.: The synchronization process for passenger-vehicle matching at taxi queue i

The effective arrival rate of vehicle is shown in equation 7.1. The derivations of vehicle incoming flows $F_{i,in}$ will be shown in the next section on network flow balance, since they are based on departure flows from all other spatial units. Similarly, we can also obtain the effective TTS vehicle arrival rate in equation 7.2.

In addition, we derive service rate μ_i^{ATS} and μ_i^{TTS}, directly from empirical observations on vehicle searching time, for instance, duration of processes e1, e2, and e3. Before figuring out the service rate measurements, we should clarify several key points. First, the $M/M/1$ queues for both service types are built at zone levels, other than taxi stands or points of interest. It may be related to zonal road network configurations and length but are less likely to be observed in reality. Second, the vehicle searching time are observable, only by counting empty trips that are fully or partially inside spatial unit i. The outside trip segments even for same vehicles are assumed to be not related to matching efficiency of the spatial unit i. Thus, under the $M/M/1$ modeling structure, we can derive service rate based on observed total system time (i.e. vehicle searching time from begins of passenger searching to pickups). In $M/M/1$, the system time follows exponential distribution, as shown in equation

7.3 and 7.4. Thus, the difference between service rate and arrival rate should be the mean system time as definitions of exponential distribution, as shown in equation 7.5 and 7.6.

$$\hat{\lambda}_i^{v,ATS} := \lambda_i^{v,ATS} + F_{i,in} p_i^{p,ATS} \tag{7.1}$$

$$\hat{\lambda}_i^{v,TTS} := \lambda_i^{v,TTS} + F_{i,in} p_i^{p,TTS} \tag{7.2}$$

$$w(t_i^{ATS}) = \left(\mu_i^{ATS} - \lambda_i^{pv,ATS}\right) e^{-\left(\mu_i^{ATS} - \lambda_i^{pv,ATS}\right)} \tag{7.3}$$

$$w(t_i^{TTS}) = \left(\mu_i^{TTS} - \lambda_i^{pv,TTS}\right) e^{-\left(\mu_i^{TTS} - \lambda_i^{pv,TTS}\right)} \tag{7.4}$$

$$\mu_i^{ATS} = \lambda_i^{pv,ATS} + \hat{t}_i^{ATS} \tag{7.5}$$

$$\mu_i^{TTS} = \lambda_i^{pv,TTS} + \hat{t}_i^{TTS} \tag{7.6}$$

where, $w(\cdot)$ is the probability density function of random variable; t_i^* is observed vehicle searching time (* can be either ATS or TTS); and $\hat{t_i^*}$ is empirical mean vehicle searching time (* can be either ATS or TTS).

7.2.3 Inclusions of Road Network Performance

Since taxicabs transport passengers through the road network, there is a close interaction between the taxi- and urban road- systems. As mentioned before, we also split the urban road system into multiple homogeneous subsystems, each of which is modeled as a $\cdot/M/c$ queue, as shown in Figure 7.3, where $1 < c < \infty$ represents

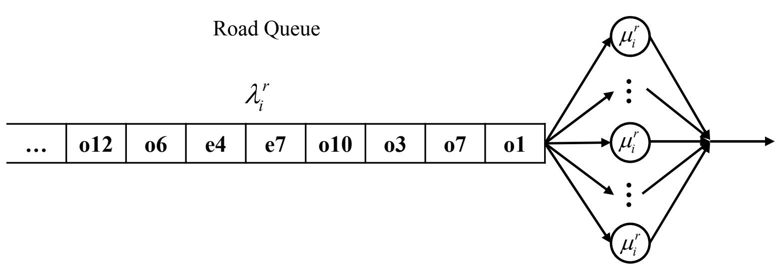


Fig. 7.3.: The queue for vehicle traveling through road queue j

the road capacity. Once a vehicle enters the road subsystem, it queues up and waits for available road space. The derivation of number of servers in each homogeneous road subsystem, c, is based on the idea of Macroscopic Fundamental Diagram (MFD) proposed and applied in recent years [114,167]. MFD models the relationship of traffic accumulation (or network density) and production (outgoing flows) and indicates a critical accumulation leading to a congested road network. The c corresponds to the critical taxi accumulations. Since the both terms reveal the maximum number of vehicles can be processed without delays. On the other hand, the derivation of service rate at each server is similar as taxi queues in equations 7.3 to 7.6, by counting vehicle travel time in one specific spatial unit and computing based on exponential distribution of observed travel time.

The last component of interest in the road queues is the effective arrival and departure flows. Since the road network does not differentiate service types and vehicle status. The effective arrivals should be a pooled flow from both the ATS and TTS containing two types of vehicle flows: (1) matched pairs (i.e. occupied vehicles transporting passengers to destinations) from taxi subsystem i, D_i^{TTS} and D_i^{ATS}; and (2) remaining vehicle arrivals in $F_{i,in}$, who just driving through the spatial unit i, regardless of searching (e.g. e7 and e4 in Figure7.1) or transporting passengers (e.g. o12 and o10 in Figure7.1). The effective arrival process is in equation 7.7. The detail analyses on the pooled flows will be presented in the next section on network flow balance. One more interesting point is about the departure flow of $M/M/c$. More complicated than occupied vehicle flow departure from $M/M/1$, the departure flow from road queue will have multiple vehicle status (occupied or empty) and service types (ATS or TTS). Considering different movement patterns, we further distribute departure flow depending on vehicle status and service types. Different type of vehicles are assigned with special routing probabilities for distribution over road network. Identification of vehicle types is primarily based on their portions in incoming flows of spatial units, which are consistent regardless of queue arrival and departure flows.

$$\lambda_i^r := D_i^{TTS} + D_i^{ATS} + \left(1 - p_i^{p,ATS} - p_i^{p,TTS}\right) F_{i,in} \tag{7.7}$$

Except for modal split between ATS and TTS, the both taxi queues interacts with each other in the ways of vehicle flow split and merges within each spatial unit, as shown in Figure7.4. Once vehicle flows enter one spatial unit, it will split based on

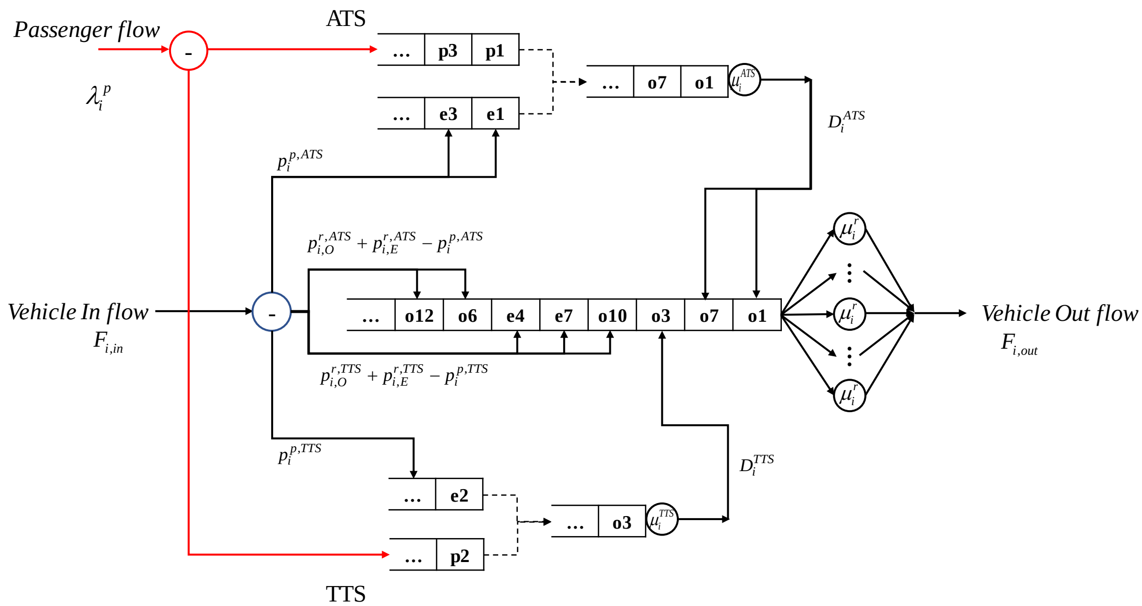


Fig. 7.4.: The subnetwork consisting of two taxi queues and one road queue at spatial unit i

service types and vehicle status. For those vehicles who can pick up new passengers in this spatial unit, $F_{i,in} p_i^{p,ATS}$, they will form external vehicle arrival for ATS queue along with newly online vehicles $\lambda_i^{v,ATS}$ and yield a departure flow from ATS queue with rate of D_i^{ATS}. Similar split is applied for TTS queue and yields a departure flow for TTS queue with rate of D_i^{TTS}. The departure flows from both taxi queues (i.e. occupied vehicles with new pickups) will queue at road queue, along with remaining vehicle arrivals of incoming flow who do not pickup any new passengers. Note that the vehicle sequence of road queue shown in Figure7.4 is just an example. Following the properties of $M/M/1$ and $M/M/c$, we can also derive following equations:

$$\lambda_i^{pv,ATS} = D_i^{ATS} \tag{7.8}$$

$$\lambda_i^{pv,TTS} = D_i^{TTS} \tag{7.9}$$

$$F_{i,out} = \lambda_i^r := D_i^{TTS} + D_i^{ATS} + \left(1 - p_i^{p,ATS} - p_i^{p,TTS}\right) F_{i,in} \tag{7.10}$$

Beyond one unique spatial unit, the incoming and outgoing flows, $F_{i,in}$ and $F_{i,out}$, can be formulated with routing probabilities as follows.

$$F_{i,in} = \sum_{j \in I} a_{ji} F_{j,out} \left(p_{ji}^{O,ATS} p_{j,O}^{r,ATS} + p_{ji}^{E,ATS} p_{j,E}^{r,ATS} + p_{ji}^{O,TTS} p_{j,O}^{r,TTS} + p_{ji}^{E,TTS} p_{j,E}^{r,TTS}\right) \tag{7.11}$$

7.3 Stationary State of Urban Taxi Queueing Network

It has been demonstrated that the joint distribution of two individual flows in the synchronization process can never reach the steady state, regardless of arrival and service rates [166]. Thus, inclusions of synchronization processes make the queueing network complicated and unstable. Therefore, this section starts from the transient and asymptotic behaviors of synchronized flows and discusses the possible approximations for synchronization processes with common queues enabling us to obtain stationary distributions for the proposed queueing network.

7.3.1 Instability of Synchronized Flows and Approximations

The proposed $SM/M/1$ queue can be presented as a two stage process by introducing a virtual buffer for synchronized pairs of passengers and vehicles, consisting of (1) virtually synchronizing two distinct flows of passengers λ_P and vehicles λ_V and immediately queueing up at a virtual buffer; and (2) serving the queued synchronized flow with a service rate of μ, as shown in the left side of Figure7.5. Let X_t^P and X_t^V be the number of passengers and vehicles in corresponding buffers at time t. $S_t = \min(X_t^P, X_t^V)$ is number of the synchronized pairs of passengers and vehicles at the virtual buffer. Prior studies [168,169] have explored the transient and asymptotic behaviors of S_t and proved that the synchronized flow S_t converges to a Poisson process both analytically and numerically. The literature further yields a $M/M/1$ approximation for $SM/M/1$, with the arrival rate $\min(\lambda^P, \lambda^V)$ and service rate $\mu > \min(\lambda^P, \lambda^V)$ shown in the right side of Figure7.5. The approximation has

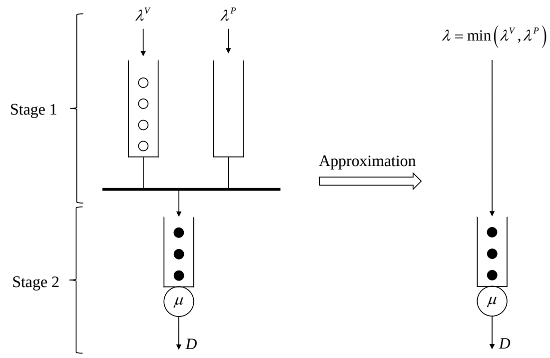


Fig. 7.5.: The general case of the $SM/M/1$ approximation with $M/M/1$

been validated by simulation demonstrating small differences in system performance metrics between the simulation and approximation, less than 1% in most cases and no more than 3% under the condition of heavy traffic (i.e. large $\min(\lambda^P, \lambda^V)/\mu$), as well as equal arrival rates of two distinct flows [168]. For completeness, we present several key properties of S_t. The details in derivations and proofs can be found in [168,169].

Assume independent Poisson arrivals of both passengers and vehicles. Observe that,

$$\begin{aligned} P(S_t = s) &= P(S_t = s \text{ and } X_t^P \geq X_t^V) + P(S_t = s \text{ and } X_t^P < X_t^V) \\ &= \sum_{x=s}^{\infty} e^{-\lambda_P t} \frac{(\lambda^P t)^x}{x!} e^{-\lambda^V t} \frac{(\lambda^V t)^s}{s!} + \sum_{y=s+1}^{\infty} e^{-\lambda_V t} \frac{(\lambda^V t)^y}{y!} e^{-\lambda^P t} \frac{(\lambda^P t)^s}{s!} \end{aligned}$$

. Obviously, this is not a Poisson process. The approximation idea is that as $t \to \infty$, S_t may approach a Poisson process so that asymptotically we have $M/M/1$-like behavior. Investigating the mean and variance of S_t, it is well-known that,

- The pair process S_t has the following asymptotic properties:

$$\begin{cases} \lim_{t\to\infty} P(S_t = X_t^P) = 1 & \text{if } \lambda^P < \lambda^V \\ \lim_{t\to\infty} P(S_t = X_t^V) = 1 & \text{if } \lambda^P > \lambda^V \\ \lim_{t\to\infty} P(S_t = X_t^V) = \lim_{t\to\infty} P(S_t = X_t^P) = \frac{1}{2} & \text{if } \lambda^P = \lambda^V \end{cases}$$

- The long-run averages of the mean and variance of the synchronization process S_t are given by

$$\begin{cases} \lim_{t\to\infty} \frac{E[S_t]}{t} = \lim_{t\to\infty} \frac{V[S_t]}{t} = \min(\lambda^P, \lambda^V) & \text{if } \lambda^P \neq \lambda^V \\ \lim_{t\to\infty} \frac{E[S_t]}{t} = \lambda^P \text{ and } \lim_{t\to\infty} \frac{V[S_t]}{t} = \lambda^P \left(1 - \frac{1}{\pi}\right) & \text{if } \lambda^P = \lambda^V \end{cases}$$

7.3.2 Network Flow Balance

It is straightforward that the queueing network can be presented by Jackson network consisting of $|I|$ $M/M/1$ queues for ATS passenger-vehicle matching, $|I|$ $M/M/1$ queues for TTS passenger-vehicle matching, and $|I|$ standard $M/M/c$ queues for road subsystems. The $M/M/1$ approximations reduce the two individual flows into a minimum one, which cannot yield a reliable track of demand and supply unbalance. However, recall the asymptotic behaviors of S_t, as $t \to \infty$, the synchronization process can match almost all arrivals in the flow with lower rate.

For each taxi subsystem $i \in I$,

$$\lambda_i^{pv,TTS} = D_i^{TTS} = \min\left(\hat{\lambda}_i^{v,TTS}, \lambda_i^{p,TTS}\right) = \min\left(\lambda_i^{v,TTS} + F_{i,in} p_i^{p,TTS}, p_i^{TTS} \lambda_i^p\right) \tag{7.12}$$

$$\lambda_i^{pv,ATS} = D_i^{ATS} = \min\left(\hat{\lambda}_i^{v,ATS}, \lambda_i^{p,ATS}\right) = \min\left(\lambda_i^{v,ATS} + F_{i,in} p_i^{p,ATS}, p_i^{ATS} \lambda_i^p\right) \tag{7.13}$$

For each road subsystem $i \in I$,

$$\begin{aligned} F_{i,out} = \lambda_i^r &= D_i^{TTS} + D_i^{ATS} + \left(1 - p_i^{p,ATS} - p_i^{p,TTS}\right) F_{i,in} \\ &= \min\left(\lambda_i^{v,TTS} + F_{i,in} p_i^{p,TTS}, p_i^{TTS} \lambda_i^p\right) + \min\left(\lambda_i^{v,ATS} + F_{i,in} p_i^{p,ATS}, p_i^{ATS} \lambda_i^p\right) \\ &\quad + \left(1 - p_i^{p,ATS} - p_i^{p,TTS}\right) F_{i,in} \end{aligned} \tag{7.14}$$

Substitute equation 7.14 into 7.11,

$$F_{i,in} = \sum_{j \in I} a_{ji}(\min(\lambda_j^{v,TTS} + F_{j,in}p_j^{p,TTS}, p_j^{TTS}\lambda_j^p) + \min(\lambda_j^{v,ATS} + F_{j,in}p_j^{p,ATS}, p_j^{ATS}\lambda_j^p)$$
$$+ (1 - p_j^{p,ATS} - p_j^{p,TTS})F_{j,in})(p_{ji}^{O,ATS}p_{j,O}^{r,ATS} + p_{ji}^{E,ATS}p_{j,E}^{r,ATS} + p_{ji}^{O,TTS}p_{j,O}^{r,TTS} + p_{ji}^{E,TTS}p_{j,E}^{r,TTS}) \quad \textbf{(7.15)}$$

The above equation 7.15 is available for all $|I|$ spatial units, which forms a system of $|I|$ equations. Solving such equation system leads to incoming flows, as well as effective arrival rate for both taxi and road queues. However, we introduce following inequalities to the equation system and convert the problem into a linear programming one, due to the existence of min sets. These inequalities are strictly consistent with min sets in equation 7.15.

$$\lambda_i^{pv,TTS} \geq \lambda_i^{v,TTS} + F_{i,in}p_i^{p,TTS} \quad \textbf{(7.16)}$$
$$\lambda_i^{pv,TTS} \geq p_i^{TTS}\lambda_i^p \quad \textbf{(7.17)}$$
$$\lambda_i^{pv,TTS} \leq \lambda_i^{v,TTS} + F_{i,in}p_i^{p,TTS} \quad \textbf{(7.18)}$$
$$\lambda_i^{pv,TTS} \leq p_i^{TTS}\lambda_i^p \quad \textbf{(7.19)}$$
$$\lambda_i^{pv,ATS} \geq \lambda_i^{v,ATS} + F_{i,in}p_i^{p,ATS} \quad \textbf{(7.20)}$$
$$\lambda_i^{pv,ATS} \geq p_i^{ATS}\lambda_i^p \quad \textbf{(7.21)}$$
$$\lambda_i^{pv,ATS} \leq \lambda_i^{v,ATS} + F_{i,in}p_i^{p,ATS} \quad \textbf{(7.22)}$$
$$\lambda_i^{pv,ATS} \leq p_i^{ATS}\lambda_i^p \quad \textbf{(7.23)}$$

Therefore, we can convert the effective arrival rate computations into a linear programming problem that can be solved in polynomial time and maintain same solutions:

$$\min_{\{\lambda_i^{pv,TTS},\lambda_i^{pv,ATS},F_{i,in}\}_{i\in I}} \sum_{i\in I} \left(\lambda_i^{pv,TTS}+\lambda_i^{pv,ATS}\right) \tag{7.24}$$

Subject to

equations 7.16to7.23, for every $i \in I$

equations 7.15, for every $i \in I$

$\lambda_i^{pv,TTS} \geq 0, \lambda_i^{pv,ATS} \geq 0, F_{i,in} \geq 0$, for every $i \in I$

7.3.3 Stationary State Distribution of Queueing Network

Recall the subnetwork in Figure7.4, the taxi system in each spatial unit behaves as an independent system of one or multiple (depending on the scale of spatial units) set of two parallel $M/M/1$ queues and one $M/M/c$ queue. We can prove the existence of a steady-state distribution for the subnetwork and derive.

Theorem 7.3.1 *If we have $\lambda_i^{pv,TTS} < \mu_i^{TTS}$, $\lambda_i^{pv,ATS} < \mu_i^{ATS}$, and $\lambda_i^r < c\mu_i^r$. Further, for the state $X = \left\{x_i^{TTS}, x_i^{ATS}, x_i^{O,TTS}, x_i^{E,TTS}, x_i^{O,ATS}, x_i^{E,ATS}\right\}$, the steady state probability is given by:*

$$\pi(X) = \begin{cases} \left(\frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right)^{x_i^{TTS}} \left(\frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right)^{x_i^{ATS}} \frac{1}{x_i!}\left(\frac{\lambda_i^r}{\mu_i^r}\right)^{x_i}\pi(\phi) & \text{if } 0 \leq x_i < c_i \\ \left(\frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right)^{x_i^{TTS}} \left(\frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right)^{x_i^{ATS}} \frac{1}{c_i^{x_i-c_i}c_i!}\left(\frac{\lambda_i^r}{\mu_i^r}\right)^{x_i}\pi(\phi) & \text{if } x_i \geq c_i \end{cases} \tag{7.25}$$

where, $x_i = x_i^{O,TTS} + x_i^{E,TTS} + x_i^{O,ATS} + x_i^{E,ATS}$,

$$\pi(\phi) = \left(1-\frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right)\left(1-\frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right)\left(\frac{\mu_i}{(c_i-1)!(c_i\mu_i^r-\lambda_i^r)}\left(\frac{\lambda_i^r}{\mu_i^r}\right)^{c_i}+\sum_{n=0}^{c_i-1}\frac{1}{n!}\left(\frac{\lambda_i^r}{\mu_i^r}^n\right)\right)^{-1}$$

Proof Following the theorem 1.13 in [170], we prove the stationary state distribution for the subnetwork as follows: Let $X(t)$ be a stationary Markov process with transition rates $q(m,n)$, where, m,n are two system states. If we can find a collection of numbers $q'(m,n)$, such that $q'(m) = q(m)$ and a collection of positive numbers $\pi(m)$, summing to unity, such that $\pi(m)q(m,n) = \pi(n)q'(n,m)$, then $q'(n,m)$ are the transition rates of the reversed process $X(\tau - t)$ and $\pi(m)$ is the equilibrium distribution of both processes.

First, given the state $m := \left\{x_i^{TTS}, x_i^{ATS}, x_i^{O,TTS}, x_i^{E,TTS}, x_i^{O,ATS}, x_i^{E,ATS}\right\}$, we can enumerate the system states and define the rates of reversed process: (1)The one TTS (similar for ATS) arrival at taxi queue i yields the state

$$n := \left\{x_i^{TTS}+1, x_i^{ATS}, x_i^{O,TTS}, x_i^{E,TTS}, x_i^{O,ATS}, x_i^{E,ATS}\right\};$$

$$q(m,n) = \lambda_i^{pv,TTS}\ , q'(n,m) = \mu_i^{TTS}.$$

(2) One occupied vehicle departing from TTS i (similar for ATS) and arriving at corresponding road queue yield the state

$$n := \left\{x_i^{TTS}-1, x_i^{ATS}, x_i^{O,TTS}+1, x_i^{E,TTS}, x_i^{O,ATS}, x_i^{E,ATS}\right\};$$

$$q(m,n) = \mu_i^{TTS}\ ,\ q'(n,m) = (x_i{+}1)\lambda_i^{pv,TTS}\mu_i^r/\lambda_i^r \text{ if } 0 \le x_i < c_i \text{ or } c_i\lambda_i^{pv,TTS}\mu_i^r/\lambda_i^r \text{ if } x_i \ge c_i.$$

(3) One vehicle departing from one road queue i and arriving at another road queue j yields the state $n := \{x_i - 1, x_j + 1\}$;

$$q(m,n) = x_i\mu_i^r p_{ij}\ , q'(n,m) = \lambda_i^r p_{ij}(x_j+1)\mu_j^r/\lambda_j^r \text{ if } 0 \le x_i < c_i \text{ and } 0 \le x_j < c_j \text{ or,}$$

$$q(m,n) = x_i\mu_i^r p_{ij}\ , q'(n,m) = \lambda_i^r p_{ij}c_j\mu_j^r/\lambda_j^r \text{ if } 0 \le x_i < c_i \text{ and } x_j \ge c_j \text{ or,}$$

$$q(m,n) = c_i\mu_i^r p_{ij}\ , q'(n,m) = \lambda_i^r p_{ij}(x_j+1)\mu_j^r/\lambda_j^r \text{ if } x_i \ge c_i \text{ and } 0 \le x_j < c_j \text{ or,}$$

$$q(m,n) = c_i\mu_i^r p_{ij}\ , q'(n,m) = \lambda_i^R p_{ij}c_j\mu_j^r/\lambda_j^r \text{ if } x_i \ge c_i \text{ and } x_j \ge c_j.$$

(4) One vehicle departing from one road queue j and arriving at one TTS (similar for ATS) queue i yields the state $n := \left\{x_i^{TTS}+1, x_i^{ATS}, x_j - 1\right\}$;

$$q(m,n) = x_j\mu_j^r p_{ji}\ , q'(n,m) = \lambda_j^r p_{ji}\mu_i^{TTS}/\lambda_i^{pv,TTS} \text{ if } 0 \le x_j < c_j \text{ or,}$$

$$q(m,n) = c_j\mu_j^r p_{ji}\ , q'(n,m) = \lambda_j^r p_{ji}\mu_i^{TTS}/\lambda_i^{pv,TTS} \text{ if } x_j \ge c_j.$$

(5) One vehicle getting destination and departing the system immediately after road queue i yields the state $n := \left\{x_i^{TTS}, x_i^{ATS}, x_i - 1\right\}$;

$$q(m,n) = x_i\mu_i^r p_{i0}\ , q'(n,m) = \lambda_i^r p_{i0} \text{ if } 0 \le x_i < c_i \text{ or,}$$

$$q(m,n) = c_i\mu_i^r\ , q'(n,m) = \lambda_i^r p_{i0} \text{ if } x_i \ge c_i.$$

With the above 5 state transitions, it can now be easily checked that for any two states i, j, the equation 7.25 is satisfied. ■

Theorem 7.3.2 *If for every taxi or road subsystem $i \in I$, we have $\lambda_i^{pv,TTS} < \mu_i^{TTS}$, $\lambda_i^{pv,ATS} < \mu_i^{ATS}$, and $\lambda_i^r < c_i \mu_i^r$. Further for the state $X = \{x_i^{TTS}, x_i^{ATS}, x_i^{O,TTS}, x_i^{E,TTS}, x_i^{O,ATS}, x_i^{E,ATS}\}_{i=1}^{|I|}$, the steady state probability is given by:*

$$\pi(X) = \begin{cases} \prod_{i\in I} \left(\frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right)^{x_i^{TTS}} \left(\frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right)^{x_i^{ATS}} \prod_{i\in I} \frac{1}{x_i!} \left(\frac{\lambda_i^r}{\mu_i^r}\right)^{x_i} \pi(\phi) & \text{if } 0 \leq x_i < c_i \\ \prod_{i\in I} \left(\frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right)^{x_i^{TTS}} \left(\frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right)^{x_i^{ATS}} \prod_{i\in I} \frac{1}{c_i^{x_i-c_i} c_i!} \left(\frac{\lambda_i^r}{\mu_i^r}\right)^{x_i} \pi(\phi) & \text{if } x_i \geq c_i \end{cases} \tag{7.26}$$

where, $x_i = x_i^{O,TTS} + x_i^{E,TTS} + x_i^{O,ATS} + x_i^{E,ATS}$,

$$\pi(\phi) = \prod_{i\in I} \left(1 - \frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}}\right) \left(1 - \frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}}\right) \prod_{i\in I} \left(\frac{\mu_i^r}{(c_i-1)!(c_i\mu_i^r - \lambda_i^r)} \left(\frac{\lambda_i^r}{\mu_i^r}\right)^{c_i} + \sum_{n=0}^{c_i-1} \frac{1}{n!} \left(\frac{\lambda_i^r}{\mu_i^r}^n\right)\right)^{-1}$$

The proof for theorem 1 is developed for the subsystem of queueing network, consisting of state transitions between taxi and road queues within one spatial unit. Then it can be easier to extend the proof to the whole queuing network, since the routing process over network is based on a fixed probability matrix or a Bernoulli splitting process. It is straightforward that the stationary state distribution of the proposed queueing network is the product of the stationary state distribution of the subsystem. This is also one classical proof in the literature, thus not presented here.

7.3.4 Performance Metrics

The $M/M/1$ and $M/M/c$ queues have well-defined performance metrics under stationary distribution. For example, we have the queue server utilization rate $\rho_i^{ATS}, \rho_i^{TTS}$, the expected number of vehicles at one queue L_i^{ATS}, L_i^{TTS}, expected number of waiting passenger-vehicle pairs $L_i^{q,ATS}$, $L_i^{q,TTS}$, expected sojourn time W_i^{ATS}, W_i^{TTS}, and expected waiting time in queue $W_i^{q,ATS}$, $W_i^{q,TTS}$ as follows:

$$\begin{aligned}
\rho_i^{ATS} &= \frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}} \ , \rho_i^{TTS} = \frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}} \\
L_i^{ATS} &= \frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS} - \lambda_i^{pv,ATS}} \ , L_i^{TTS} = \frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS} - \lambda_i^{pv,TTS}} \\
L_i^{q,ATS} &= \frac{(\lambda_i^{pv,ATS})^2}{\mu_i^{ATS}(\mu_i^{ATS} - \lambda_i^{pv,ATS})} \ , L_i^{q,TTS} = \frac{(\lambda_i^{pv,TTS})^2}{\mu_i^{TTS}(\mu_i^{TTS} - \lambda_i^{pv,TTS})} \\
W_i^{ATS} &= \frac{1}{\mu_i^{ATS} - \lambda_i^{pv,ATS}} \ , W_i^{TTS} = \frac{1}{\mu_i^{TTS} - \lambda_i^{pv,TTS}} \\
W_i^{q,ATS} &= \frac{\lambda_i^{pv,ATS}}{\mu_i^{ATS}(\mu_i^{ATS} - \lambda_i^{pv,ATS})} \ , W_i^{q,TTS} = \frac{\lambda_i^{pv,TTS}}{\mu_i^{TTS}(\mu_i^{TTS} - \lambda_i^{pv,TTS})}
\end{aligned}$$

For each road queue $i \in I$, given λ_i^r and μ_i^r, we can also derive similar system performance metrics as follows:

$$\begin{aligned}
\rho_i^r &= \frac{\lambda_i^r}{c_i \mu_i^r} \ , L_i^r = c_i \rho_i^r + \left(\frac{(c_i \rho_i^r)^{c_i} \rho_i^r}{c_i!(1-\rho_i^r)^2} \right) p_i(0) \ , L_i^{q,r} = \left(\frac{(c_i \rho_i^r)^{c_i} \rho_i^r}{c_i!(1-\rho_i^r)^2} \right) p_i(0) \\
W_i^r &= \frac{1}{\mu_i^r} + \left(\frac{(c_i \rho_i^r)^{c_i} \rho_i^r}{c_i! c_i \mu_i^r (1-\rho_i^r)^2} \right) p_i(0) \ , W_i^{q,r} = \left(\frac{(c_i \rho_i^r)^{c_i} \rho_i^k}{c_i! c_i \mu_i^r (1-\rho_i^r)^2} \right) p_i(0)
\end{aligned}$$

where, ρ_i^r is the utilization rate of road queue, L_i^r is the expected queue length of the system, $L_i^{q,r}$ is the expected queue length waiting for service; W_i^r is the expected sojourn time; and $W_i^{q,r}$ is the expected waiting time before service begins; and $p_i(0)$ is the probability of empty queues, derived from $\left(\frac{r^{c_k}}{c_k!(1-\rho) + \sum_{n=0}^{c_k-1} \frac{r^n}{n!}} \right)^{-1}$.

Within the proposed queueing network consisting of taxi queues $M/M/1$ and road queues $M/M/c$, we have average number of vehicles in network L_I, total average load on network γ_I, and average delay throughout the network W_I.

$$L_I = \sum_{i \in I} L_i^r + \sum_{i \in I} (L_i^{ATS} + L_i^{TTS}) \ , \gamma_I = \sum_{i \in I} \left(\hat{\lambda}_i^{v,TTS} + \hat{\lambda}_i^{v,ATS} \right) \ , W_I = \frac{L_I}{\gamma_I}$$

7.4 Case Study

7.4.1 The Case of New York City

In this section, we apply our proposed queueing network into the competitive taxi market of New York City (NYC), where has one of the largest TTS (fleet size of more than 13,000 yellow taxicabs) and ATS market (weekly active Uber drivers more than 45,000) in North America, as of April 2017. The case study is developed with multiple datasets: (a) the ride records of both yellow taxicabs and for-hire vehicle (Uber) shared by the NYC Taxi & Limousine Commission (ref:`http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml`), containing a time-stamped trip record of 6-tuples (O, D, t, tt, d, F), which denote the geolocation of origin and destination, pickup time, in-vehicle travel time, trip distance, and total fares respectively; (b) the Uber trajectory and operation data directly crawled from Uber platform with high-frequency (every 5s) empty vehicle trajectory information, as well as surge pricing and estimated waiting time at 470 specific locations [?]; (c) the city traffic flow data with 15-minute short count link volume and speed data in one specific week of each year, shared by New York Department of Transportation (ref:`https://www.dot.ny.gov/divisions/engineering/technical-services/highway-data-services/hdsb`); and (d) the Google Maps Directions API for shortest route planning between the pair of locations (ref:`https://developers.google.com/maps/documentation/directions/intro`).

This study also introduces the idea of a 'homogeneous region', which is defined based on the Poisson assumption on the passenger and vehicle arrival process. We perform extensive hypothesis tests of the Poisson assumption under different spatial scales (e.g. Borough[1], community districts[2], zip code tabulation area [ZCTA] [3], and

[1] 5 in total and each is with an average area of 60.4 mi^2, see `https://en.wikipedia.org/wiki/Boroughs_of_New_York_City`

[2] 71 in total if we include regions of airports and parks, and each is with an average of 4.3mi^2, see `https://www1.nyc.gov/site/planning/community/community-portal.page`

[3] 214 in total, and each is with an average of 1.4 mi^2, see `https://www.census.gov/geo/reference/zctas.html`

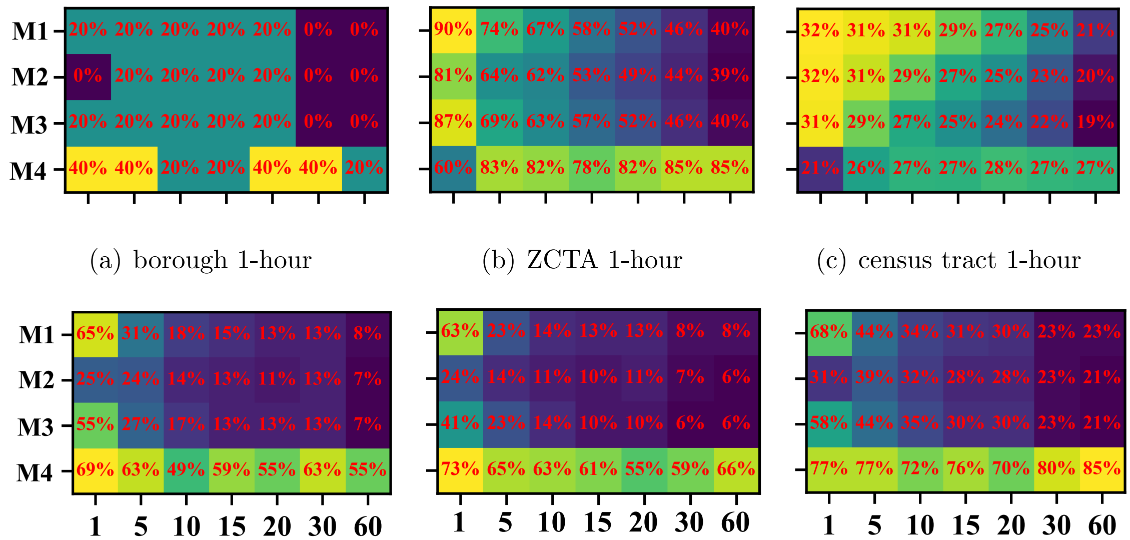
census tract [4]), as well as arrival count interval (varying from 1 minute to 1 hour) and study period (time-of-the-day and day-of-the-week). The Kolmogorov Smirnov (KS) test is adapted for discrete distribution [149] and is used to test whether the passenger or vehicle arrivals can be assumed to be Poisson distributed. In addition, three additional χ^2 distribution based tests [171] (e.g. Anscombe, Likelihood, Conditional) are applied to test whether the arrivals follow a Poisson distribution. The key test results for passengers and vehicle arrivals during peak hours are summarized in Figure 7.6 and 7.7.. Note that we skip several results due to space limits and share full results as one online supplement file[5]. To sum up, we have the following findings for case study development: a) the study period should be limited to one hour peak (from 6 to 7pm) or off peak (from 10 to 11 am) of every Mondays to Thursdays, which leads to more spatial units holding Poisson arrival assumptions; and b) we should aggregate trip records at community districts (71 in total, $\sim$4.3 mi^2 on average per community district) and 1-minute count interval, which has higher probability of being in line with our Poisson arrival flow assumptions. Note that we skip several results due to space limits. To sum up, we have following findings for case study development: a) the study period should be limited to one hour peak (from 6 to 7pm) or off peak (from 10 to 11 am) of every Mondays to Thursdays, which leads to more spatial units holding Poisson arrival assumptions; and b) we should aggregate trip records at community districts (71 in total, $\sim$4.3 mi^2 on average per community district) and 1-minute count interval, which has higher probability of being in line with our Poisson arrival flow assumptions.

7.4.2 Queue Inputs

Under proposed spatiotemporal aggregation scales, we further investigate the passenger and vehicle arrival rates $\lambda_i^p, \lambda_i^{v,ATS}, \lambda_i^{v,TTS}$. Figure 7.8 (a) to (c) exhibits the

[4] 2165 in total, and each is with an average of 0.14 mi^2, see `https://www.census.gov/geo/reference/gtc/gtc_ct.html`

[5] see `https://github.com/wenbo-purdue-git/isttt-23-taxi-system-modeling-`

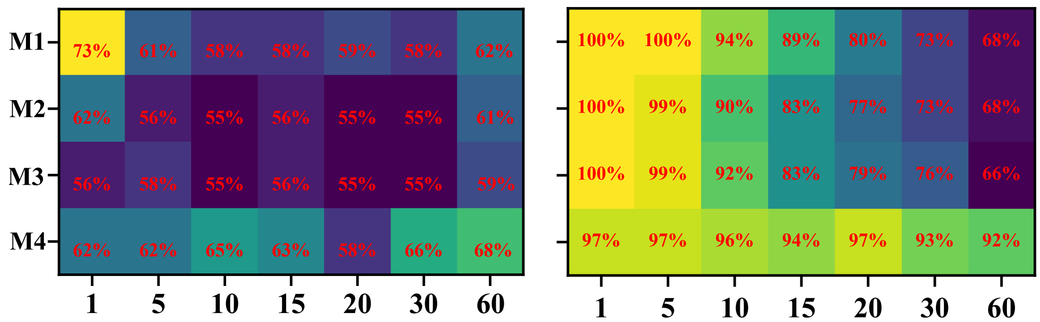


(a) borough 1-hour (b) ZCTA 1-hour (c) census tract 1-hour

(d) community districts 3-hour (e) community districts all days (f) community districts 1-hour

Fig. 7.6.: The percentage of community districts with significant hypothesis testing results under various scenarios: (a) borough and 1-hour peak of Mondays to Thursdays; (b) zcta and 1-hour peak of Mondays to Thursdays; (c) census tract and 1-hour peak of Mondays to Thursdays; (d) community districts and 3-hour peak of Mondays to Thursdays; (e) community districts and 1-hour peak of all days; and (f) community districts and 1-hour peak of Mondays to Thursdays. Moreover, x-axis indicates the arrival count interval (minutes), and y-axis denotes test methods (M1: Anscombe, M2: Likelihood, M3: Conditional, and M4: adapted KS).

p values from hypothesis testing in each community district. The red color indicates small p values less than 0.05, which reject the null hypothesis of Poisson arrivals at confidence level of 95%. It is straightforward that most community districts can be assumed to have Poisson passenger and vehicle arrivals. Figure 7.8 (d) to (f) show corresponding arrival rates. Downtown Manhattan areas have relatively higher passenger arrival rates more than 100 passengers per minute. In contrast, either ATS or TTS vehicle arrival rates are less than 5 vehicles per minute and do not have many

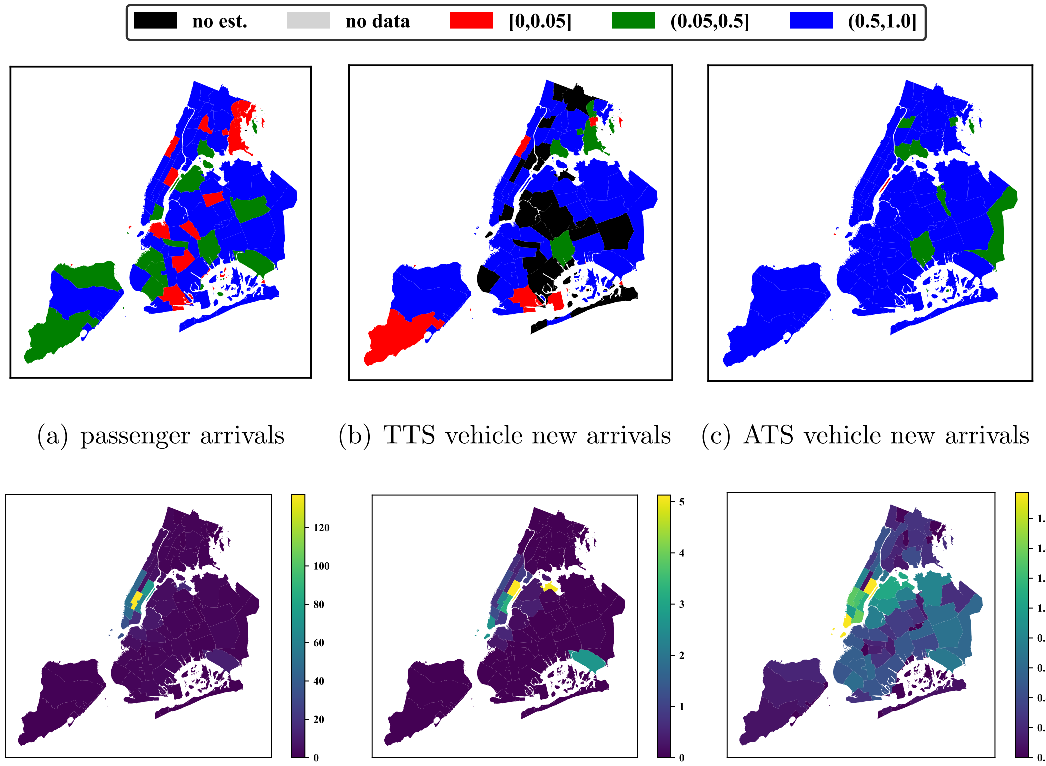


(a) TTS vehicle arrivals (b) ATS vehicle arrivals

Fig. 7.7.: The percentage of community districts with significant hypothesis testing results. X-axis indicates the arrival count interval (minutes), and y-axis denotes test methods (M1: Anscombe, M2: Likelihood, M3: Conditional, and M4: adapted KS).

variants across space, since the vehicle arrival only counts new onlines and excludes those who have been in system and move from other community districts.

Except for external passenger and vehicle arrivals, we also examine the fixed probabilities of modal split p_i^{ATS}, pickup $p_i^{p,ATS}$or$p_i^{p,TTS}$, and system-exit $p_{i0}^{E,ATS}$or$p_{i0}^{O,TTS}$. Figure 7.9 (a) to (e) present mean value of corresponding probabilities across minutes. Figure 7.9 (f) to (j) summarize variance value of corresponding probabilities across minutes. The dark color indicates almost zero variance, which provides strong empirical evidence of fixed probabilities. We can take mean value of corresponding probabilities as mode inputs. Regarding empirical observations on vehicle passenger searching time, t_i^{ATS} can be directly counted from trajectory dataset. However, the measurement on t_i^{TTS} is relatively more complicated, by introducing shortest path planning data from Google API. We segment TTS rides with Google shortest path then enumerate defined passenger searching time or travel time. Also, the vehicle routing matrix is derived based on similar measurement method as t_i^{ATS} and t_i^{TTS}. The last input of interest is the number of servers at road queues, c_i. Here, the service rate can also be defined as the inverse of expected travel time under free flow in each road subsystem. We employ the free flow speed and together with measured



(a) passenger arrivals (b) TTS vehicle new arrivals (c) ATS vehicle new arrivals

(d) λ_i^p (e) $\lambda_i^{v,TTS}$ (f) $\lambda_i^{v,ATS}$

Fig. 7.8.: The hypothesis test p value (subplots a to c) and arrival rate (subplots d to f) of both passenger and new vehicles during peak hours

trip distance by datasets and estimated trip distance from segmented shortest paths, obtain expected travel time under free flow. The number of servers are mainly generated based on the critical taxi accumulation in the MFD-like shapes, as shown in Figure7.10. Each borough tends to have consistent critical accumulations and has small differences from others.

7.4.3 Model Evaluation

In this section, we examine the model performance based on proposed settings and solutions of the linear programming problem 7.24. First, we compare the estimated $\lambda_i^{pv,*}$ that denotes the paired flow arrival rate in the synchronization process, to the observed passenger pickup rates in reality. Figure 7.11 shows almost same patterns between peak and off peak hours, but reveals many differences between ATS and TTS. The ATS system presents much lower absolute percentage errors (i.e. <5%) in almost every spatial unit. In contrast, the proposed modeling structure has reliable outputs for "hot" areas of TTS system, where attract more than 90% TTS activities in reality. Such significant differences may arise from spatial distribution of both services. Since the modeling structures involves vehicle movement over road network and routing probabilities, which directs majority of vehicles to "hot" areas and leads to unreliable estimations for remaining areas. In addition, the low percentage error also provides strong empirical evidence of $SM/M/1$ approximation with $M/M/1$.

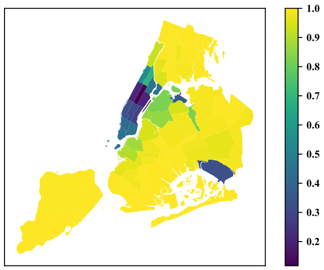
Moreover, we examine the taxi system performance, in terms of sojourn time (or total time from arrival to departure in one specific queue system), and evaluate the accuracy for both ATS and TTS, summarized in Figure 7.12. Note that, multiple black spatial units in Figure 7.12 are without accuracy measurement, since there are very limited empirical observations, resulting in unreliable measurements. Overall, the model perform well for ATS system and relatively worse for TTS system, which is in line with findings in $\lambda_i^{pv,*}$. For most spatial units, the proposed modeling structures can yield accurate measurements on ATS vehicle-passenger matching performance with less than 10% relative errors. However, same as $\lambda_i^{pv,*}$ estimation, the outer Manhattan areas (gray areas in Figure 7.12) still have worse outputs of TTS system performance. The reason is attributed to the limited number of TTS trips outside Manhattan and the unbalanced distribution of these trips.

Other performance metrics can be similarly examined at either taxi or road queues. We will not present all results due to the space limitations. Overall, the proposed queueing network model can describe vehicle-passenger matching process well and yields reliable estimations on both the expected and distribution of matching performance (e.g. sojourn time). At appropriate spatial and temporal aggregation (such as homogeneous spatial units and periods in this study), vehicles and passengers behave identically without much fewer variants, which is also our assumptions for the queueing network. However, the queueing network performance relies on the parameter calibrations that requires high-resolution dataset of individual movements. Our datasets are now are not of sufficient resolution to allow the calibration of all parameters (for example, TTS in outer Manhattan areas).

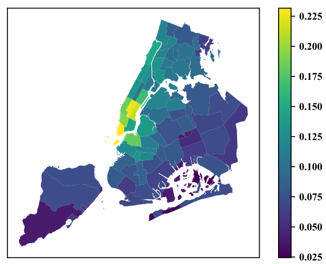
7.5 Conclusions and Future Works

The study develops a queueing network approach to describe the complex interactions between the ATS and TTS systems within one unified oligopolistic taxi market, as well as between the taxi- and urban road- system. We first introduce the queueing network structures in which not only the queue node itself can capture the dynamics of taxi passenger and vehicle behaviors but also the node connections can allow the flow exchanges accounting for network externalities. Specifically, we propose (1) the synchronized process $SM/M/1$ for both ATS and TTS passenger-vehicle matching behaviors; (2) the multi-server $M/M/c$ queue for the urban road system; and (3) the state-dependent service rate of $SM/M/1$. Moreover, we provide an approximation on the proposed non-stationary queueing network with a Jackson network and investigate the stationary state distributions under the approximations. Finally, we fully utilize our rich dataset, provide numerical examples of the NYC oligopolistic taxi market, and examine the proposed queueing network approach.

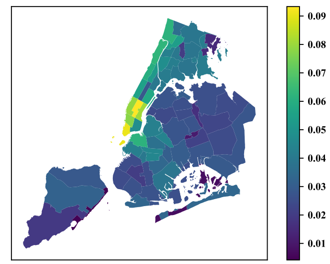
Overall, the application of the proposed modeling structure is far beyond what we have presented and illustrated. One main characteristic of ATS is the dynamic pricing and thus the drivers' and passengers' incentives. The differentiated controls over the both space and time make the problem more interesting but also challenging which are not completely addressed in this study. The model in this paper emphasizes the macroscopic interactions between urban road and taxi systems, but does not capture the behavioral dynamics of the individuals and how they respond to the taxi market. In future studies, the proposed queueing network will be generalized to include the full dynamics of taxi markets and individual behaviors, thus allowing us to have in-depth insights into system control. Further, sensitivity analysis of the model should also be conducted to understand the stability of the outputs to minor changes in taxi supply and demand.



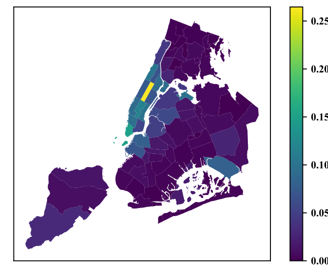
(a) mean of p_i^{ATS}



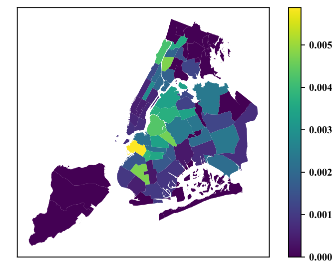
(b) mean of $p_i^{p,ATS}$



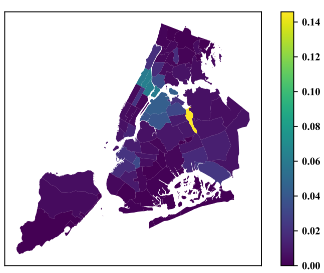
(c) mean of $p_{i0}^{E,ATS}$



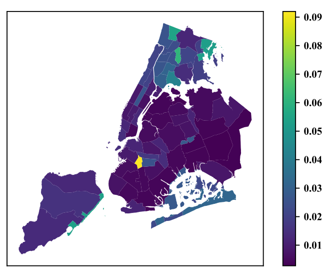
(d) mean of $p_i^{p,TTS}$



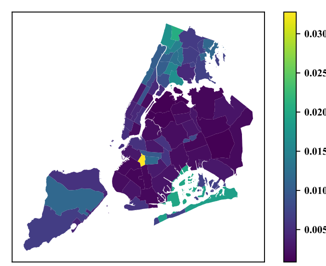
(e) mean of $p_{i0}^{O,TTS}$



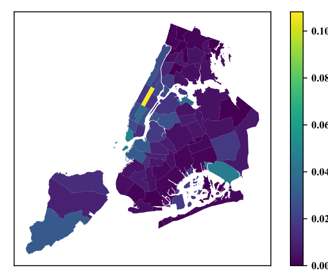
(f) variance of p_i^{ATS}



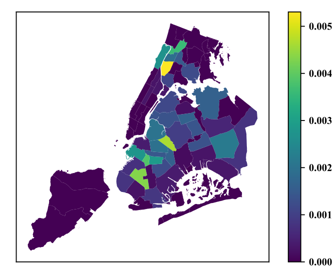
(g) variance of $p_i^{p,ATS}$



(h) variance of $p_{i0}^{E,ATS}$



(i) variance of $p_i^{p,TTS}$



(j) variance of $p_{i0}^{O,TTS}$

Fig. 7.9.: The mean and variance of modal split, pickup, and system-exiting probabilities during peak hours

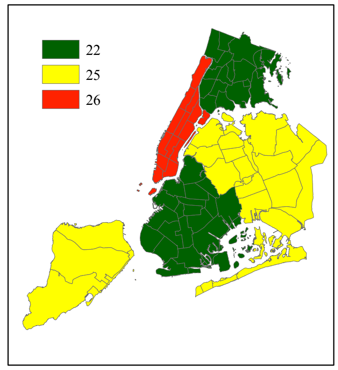
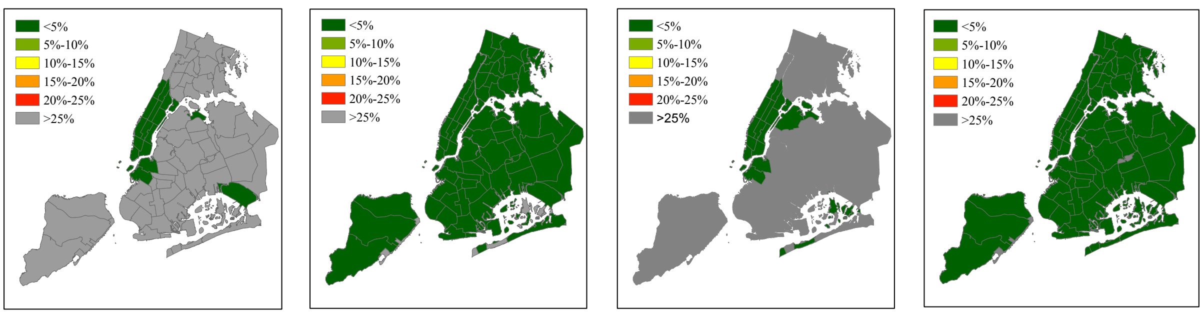
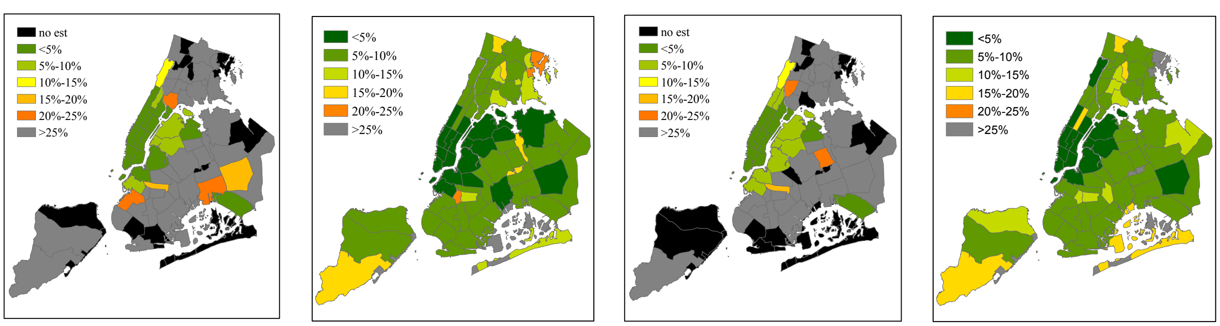


Fig. 7.10.: The number of servers of $M/M/c$ road queues



(a) TTS - peak hours (b) ATS - peak hours (c) TTS - off peak hours (d) ATS - off peak hours

Fig. 7.11.: The absolute percentage errors between estimated $\lambda_i^{pv,*}$ and observed passenger pickup flows



(a) TTS - peak hours (b) ATS - peak hours (c) TTS - off peak hours (d) ATS - off peak hours

Fig. 7.12.: The mean absolute percentage errors between expected sojourn time and observed one at taxi queues

8. CONCLUSIONS

8.1 General Findings

This dissertation explores the fundamental problems arising from disruptive mobility - app-based taxi services. The innovative urban mobility service leverages smartphone based apps and has been expanding explosively. However, we are still lacking a better understanding of such popular system and are experiencing challenges driven by three stakeholders: (i) passenger generation and competition, (ii) flexible ATS supply, and (iii) real-time fare regulation by the platform. This dissertation is motivated by the high demand for ATS system investigations under real-time control, as well as the complex interactions with external systems. Our proposed modeling structures provide insights into how demand, supply, and stakeholders interact within one large-scale system, but also how they interplay with traditional taxi and urban road systems. These studies are valuable for system evaluation, strategy development, transportation planning, and policymaking. Furthermore, new paradigms are developed to fill current gaps in the literature.

The taxi system, regardless of whether it is an ATS or TTS, is a typical two-sided platform with passengers and vehicles, involving multiple fundamental processes including demand generation, supply dynamics, demand-supply matching, and movement over the urban road network. However, understanding the details of all these fundamental processes is challenging, in particular in a large-scale system/city with limited data availability. This dissertation addresses various challenges and first narrows the gaps in high-resolution mobility data acquisition for the ATS. A novel data collection scheme provides us with a detailed trajectory and real-time control information, which builds a solid empirical background in behavioral modeling and system evaluations.

Understanding demand and supply is crucial and necessary for further system explorations. The literature has made extensive progress on demand generation but is limited to one unique taxi service. Moreover, there has been less attentions paid to the interactions among demand, supply, and dynamic pricing. In specifics, what the impacts of dynamic pricing are on demand and supply, and vice versa. This dissertation primarily examines such complex interactions. From empirical analyses, we confirm that dynamic pricing is designed based on demand and supply, but also has additional unknown impacts or rules. Inversely, dynamic pricing can be effective in controlling demand generation but is useless for the supply side, regardless of whether driver enter or exit the ATS system. On the other hand, one large-scale taxi system is more likely to exhibit complicated spatial and temporal relationship, as well as modal correlations among similar services. We also empirically validate the proposed appropriate modeling structure to address these concerns.

The more complicated process within the taxi system is vehicle-passenger matching. Although various modeling structures have been developed for the taxi system, ranging from equilibrium to agent based simulation, the descriptions of vehicle-passenger matching are too simple or at an aggregated level without considerations over stochasticity in the process. This study aggregates demand and supply arrival flows at appropriate scales for Poisson assumptions, and proposes queue theoretic approaches to capture such stochastic process. Based on our investigations, we can describe vehicle-passenger matching in the majority of spatial units with simple $M/M/1$, under appropriate aggregation scales, such as 1 minute count intervals and on average a 4.3 mi^2 spatial unit. One more challenging point is the interaction between taxi system and the urban road network, which is ignored in most studies. We also introduce a queue theoretic approach for vehicle movement in spatial units and derive queueing network to represent network externalities. Addressing both the stochasticity within the taxi system and interactions with the urban road network makes our modeling structure more reliable and unique while quantifying taxi service performance in a large-scale system.

8.2 Contributions

This dissertation improves current understandings of the app-based taxi system in a large-scale system. In addition to the aforementioned general findings, we also summarize contributions related to the designed objectives.

- *Collect a large-scale mobility dataset for one representative ATS platform that contains high-resolution information on vehicle and passenger movements, as well as real-time fare rate.*

  This objective is achieved in Chapter 3. This chapter presents the comprehensive high-resolution mobility data acquisition approach, ranging from scrape tool design and data collection scheme development to data preprocessing. The data acquisition method is novel and cost-saving for one large-scale system.

- *Characterize the spatiotemporal variations in passenger and empty vehicle flows of the ATS versus TTS at the aggregate level. The specific questions relating to the objective are listed as follows:*

  Q1: *Are the released dataset sufficient for solving our problem? If not, how can we obtain TTS and ATS operation data?*

  Q2: *What are the appropriate spatial and temporal scales for our problem?*

  The objective, along with two questions, are also completed in Chapter 3. We employ various hypothesis testing methods to validate Poisson assumptions of both the demand and supply generation processes. Alongside appropriate spatial and temporal aggregation scales, we also explore the time-of-day and day-of-the-week impacts on homogeneous generation. To our knowledge, this is one of first studies to undertake such analyses.

- *Develop a demand and supply generation model for ATS and TTS taking into account surge pricing. The specific questions relating to the objective are listed as follows*

Q3: *What are the impacts of dynamic pricing on ATS, TTS, or both ridership?*

Q4: *What exogenous variables are influential for demand generation?*

Q5: *What are the impacts of dynamic pricing, as well as exogenous variables, on ATS supply generation?*

Q6: *What modeling structures are appropriate for such spatiotemporal dataset?*

Q7: *Are there correlations occurring while the ATS and TTS rides generate? If so, how should we adjust the modeling structure?*

This objective is primarily met in Chapter 4. We fill gaps in a two-fold process. First, we combine both exogenous and endogenous variables to describe such generation processes. In particular, we identify the impacts of dynamic pricing. Second, a mixture model of spatial lag and Poisson regression is verified with better specifications, which can address spatial autocorrelation well. An additional contribution to modeling structures is about the modal correlations. Our empirical findings indicate that specifications for summations of both the ATS and TTS demand perform better than separate specifications for each demand and introduction of correlated terms.

- *Develop a decision model for ATS drivers that can describe their aggregated platform-exiting behaviors.*

  Q8: *What internal factors contribute to drivers' choices in exiting ATS platform?*

  Q9: *Are there differences in exit probabilities across aggregated zones? If yes, what variables are relevant?*

  Q10: *What modeling structures are appropriate for modeling such aggregated decisions?*

  This objective is fulfilled in Chapter 5. Instead of addressing individual driver platform-exiting behaviors, an aggregated analysis is developed with a mixed model structure of spatial lag and one-way fixed panel analysis. In this study, we also confirm the spatial autocorrelation in platform-exiting behaviors, which can be seen as a common characteristic of taxi activities at the aggregated level. In addition, the model estimations reveal spatial heterogeneity in platform-exiting behaviors. The most surprising finding is that the aggregated behaviors are not responsive to incentives and high demand on the ATS platform. This is not in line with limited existing findings at the individual level.

- *Develop a model to investigate the dynamic pricing generation considering demand, supply, and potential spatio-temporal effects.*

  Q11: *What internal factors contribute to dynamic pricing generation?*

  Q12: *Does spatio-temporal heterogeneity exist while generating dynamic pricing?*

  Q13: *Which modeling structures perform better?*

  This objective is addressed in Chapter 6, by proposing various multi-class classification methods. We also identify a better method based on various performance metrics. Moreover, we confirm the impacts of demand, supply, and spatio-temporal heterogeneity, while developing a surge pricing multiplier. However, the average multiplier over a relatively larger spatial unit (compared to the Uber defined small hexagon) will lead to an unbalanced data structure with a majority of 1.0. Very limited methods and metrics are adapted for such circumstances.

- *Devise a unified modeling framework that not only describes the stochastic nature of passenger-vehicle matching, but also captures the traffic state on the urban road network.*

  Q14: *How do drivers meet passengers within an aggregated region?*

  Q15: *How do taxi vehicle movements interact with regular traffic flow over the urban road network?*

This objective is carried out in Chapter 7. The vehicle-passenger matching and taxi movement over the urban road network are described with $M/M/1$ and $M/M/c$ queues, respectively, to capture dynamics within specific spatial units. The network of all queues formulate flow transfers among spatial units and capture network externalities. Under the queueing network, we validate not only the assumption of the fixed Bernoulli splitting process and Poisson arrival flows, but also the stationary state of the queueing network. The unified modeling structure provides a product form solution, thus is efficient with a large scale system.

8.3 Future Research

Multiple unsolved or challenging points are summarized here as extensions of the dissertation.

- The potential adaption for multi-class classification approaches while modeling unbalanced categories. As stated in the corresponding chapter, one big concern over surge pricing multiplier classification is the unbalanced data structure with a majority of 1.0. Almost all proposed predictive classifiers are inferior in yielding a reliable classification for multipliers other than 1.0. The modeling structure should have further corrections for such dataset.

- The specialty of airports. While modeling demand and supply, we prefer separate analyses for regions of interests. Airports are also defined as special regions (remote but different from neighboring regions) and require the development of separate model. However, when limited to location attributes, it is relatively difficult to enhance model performance. In future, a special analysis is expected to be implemented.

- Matching queues for vehicles and passengers arrival. In current modeling, the matching queue is approximated by the minimum number of vehicle or passenger arrival flows, which can reduce the synchronization process to a typical $M/M/1$ model. However, the matching process can undertake further exploration of randomness and stochasticity, in particular while modeling homogeneous spatial units rather than fixed stands or stations.

- Pointwise stationary distribution of queueing network for heterogeneous arrivals. The queuing network is proposed under homogeneous periods with constant arrival rates and Bernoulli splitting behaviors. However, it can extend to the case of heterogeneous periods with different rates and behaviors. A potential solution is pointwise stationary distribution, which employs a product form of stationary state distribution in every sub-homogeneous period.

- State-dependent extension to the queueing network. The state-dependent queueing network will introduce more benefits. In specifics, the inclusions of dynamic pricing into the queueing network can result in deep investigations of control strategies. However, this will violate the product-form stationary state distribution and transform into a new game theoretic equilibrium. The complexity of the state-dependent queueing network is challenging but also interesting.

REFERENCES

REFERENCES

[1] Department of Economic Population Division and United Nations Social Affairs. World urbanization prospects, the 2014 revision. Technical report, https://esa.un.org/unpd/wup/, 2014.

[2] Francois-Joseph Van Audenhove, Oleksii Korniichuk, Laurent Dauby, and Jerome Pourbaix. The future of urban mobility 2.0. Technical report, http://www.uitp.org/sites/default/files/members/140124%20Arthur%20D.%20Little%20%26%20UITP_Future%20of%20Urban%20Mobility%202%200_Full%20study.pdf, 2014.

[3] Ashley Z. Hand. Urban mobility in a digital age. Technical report, https://static1.squarespace.com/static/57c864609f74567457be9b71/t/57c9059b9de4bb1598eeee49/1472793280502/Transportation+Technology+Strategy_2016.pdf, 2016.

[4] Susan A. Shaheen, Nelson D. Chan, and Teresa Gaynor. Casual carpooling in the san francisco bay area: Understanding user characteristics, behaviors, and motivations. *Transport Policy*, 51:165–173, 2016.

[5] Lisa Rayle, Danielle Dai, Nelson Chan, Robert Cervero, and Susan Shaheen. Just a better taxi? a survey-based comparison of taxis, transit, and ridesourcing services in san francisco. *Transport Policy*, 45(Supplement C):168 – 178, 2016.

[6] Jonathan V. Hall and Alan B. Krueger. An analysis of the labor market for uber's driver-partners in the united states. *ILR Review*, page 0019793917717222, 2017.

[7] City of New York Office of Mayor. For-hire vehicle transportation study. Technical report, http://www1.nyc.gov/assets/operations/downloads/pdf/For-Hire-Vehicle-Transportation-Study.pdf, 2016.

[8] Statista. Number of smartphone users worldwide from 2014 to 2020 (in billions), June 2016.

[9] Lee Rainie and Andrew Perrin. 10 facts about smartphones as the iphone turns 10, June 2017.

[10] Joe P. Pavelka. Smartphone use during travel. Technical report, http://scholarworks.umass.edu/ttracanada_2016_conference/6/, May 2016.

[11] AAron Smith. Shared, collaborative and on demand: The new digital economy. Technical report, http://www.pewinternet.org/2016/05/19/on-demand-ride-hailing-apps/, May 2016.

[12] Hai Yang, Yan Wing Lau, Sze Chun Wong, and Hong Kam Lo. A macroscopic taxi model for passenger demand, taxi utilization and level of services. *Transportation*, 27(3):317–340, Jun 2000.

[13] Chen Mu and Xiangmo Zhao. Dynamic passenger od distribution and system performance of taxi operation system. *International Journal of Information Engineering and Electronic Business(IJIEEB)*, 2011.

[14] K. I. Wong, S. C. Wong, M. G. H. Bell, and Hai Yang. Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing markov chain approach. *Journal of Advanced Transportation*, 39(1):81–104, 2005.

[15] Jianmin Xu, S. C. Wong, Hai Yang, and Chung-On Tong. Modeling level of urban taxi services using neural network. *Journal of Transportation Engineering*, 125(3):216–223, 1999.

[16] K.I. Wong, S.C. Wong, Hai Yang, and J.H. Wu. Modeling urban taxi services with multiple user classes and vehicle modes. *Transportation Research Part B: Methodological*, 42(10):985 – 1007, 2008.

[17] Hai Yang, Cowina W.Y. Leung, S.C. Wong, and Michael G.H. Bell. Equilibria of bilateral taxi–customer searching and meeting on networks. *Transportation Research Part B: Methodological*, 44(8):1067 – 1083, 2010.

[18] Jian Wang, Guanglin Sun, and Xiaowei Hu. Game theory-based optimization of taxi fare in taxi market of china. In *Proceedings of 93rd Annual Meeting of the Transportation Research Board*, January 2014.

[19] Hai Yang, Min Ye, Wilson Hon-Chung Tang, and Sze Chun Wong. A multi-period dynamic model of taxi services with endogenous service intensity. *Operations Research*, 53(3):501–515, 2005.

[20] Nicholas Buchholz. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. Technical report, Princeton University, 2016.

[21] Guillaume R Fréchette, Alessandro Lizzeri, and Tobias Salz. Frictions in a competitive, regulated market: Evidence from taxis. Working Paper 24921, National Bureau of Economic Research, August 2018.

[22] Xianyuan Zhan, Xinwu Qian, and Satish V. Ukkusuri. A graph-based approach to measuring the efficiency of an urban taxi service system. *IEEE Transactions on Intelligent Transportation Systems*, 17(9):2479–2489, Sept 2016.

[23] Michal Maciejewski and Joschka Bischoff. Large-scale microscopic simulation of taxi services. *Procedia Computer Science*, 52(Supplement C):358 – 364, 2015. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).

[24] Fang He and Zuo-Jun Max Shen. Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C: Emerging Technologies*, 58(Part A):93 – 106, 2015.

[25] Xiaolei Wang, Fang He, Hai Yang, and H. Oliver Gao. Pricing strategies for a taxi-hailing platform. *Transportation Research Part E: Logistics and Transportation Review*, 93(Supplement C):212 – 231, 2016.

[26] Xinwu Qian and Satish V. Ukkusuri. Taxi market equilibrium with third-party hailing service. *Transportation Research Part B: Methodological*, 100(Supplement C):43 – 63, 2017.

[27] Wenbo Zhang, Satish V. Ukkusuri, and Jian John Lu. Impacts of urban built environment on empty taxi trips using limited geolocation data. *Transportation*, 44(6):1445–1473, Nov 2017.

[28] Won Kyung Lee and So Young Sohn. Taxi vacancy duration: a regression analysis. *Transportation Planning and Technology*, 40(7):771–795, 2017.

[29] Ci Yang and Eric J. Gonzales. Modeling taxi trip demand by time of day in new york city. *Transportation Research Record: Journal of the Transportation Research Board*, 2429:110 – 120, 2014.

[30] Wangqi Ge, Dan Shao, Meigen Xue, Hong Zhu, and Jie Cheng. Urban taxi ridership analysis in the emerging metropolis: Case study in shanghai. *Transportation Research Procedia*, 25(Supplement C):4916 – 4927, 2017. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016.

[31] Wenbo Zhang, Xinwu Qian, and Satish V. Ukkusuri. *Identifying the Temporal Characteristics of Intra-City Movement Using Taxi Geo-Location Data*, pages 68–88. IGI Global, Hershey, PA, 2017.

[32] Xinwu Qian and Satish V. Ukkusuri. Spatial variation of the urban taxi ridership using gps data. *Applied Geography*, 59(Supplement C):31 – 42, 2015.

[33] Daisik Nam, Kyung (Kate) Hyun, Hyunmyung Kim, Kijung Ahn, and R. Jayakrishnan. Analysis of grid cell–based taxi ridership with large-scale gps data. *Transportation Research Record: Journal of the Transportation Research Board*, 2544:131–140, 2016.

[34] Bruce Schaller. A regression model of the number of taxicabs in u.s. cities. *Journal of Public Transportation*, 8(5):63–78, 2005.

[35] Wai Yuen Szeto, Ryan Cheuk Pong Wong, Sze Chun Wong, and Hai Yang. A time-dependent logit-based taxi customer-search model. *International Journal of Urban Sciences*, 17(2):184–198, 2013.

[36] R.C.P. Wong, W.Y. Szeto, and S.C. Wong. Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications. *Transport Policy*, 33(Supplement C):73 – 81, 2014.

[37] Jinjun Tang, Han Jiang, Zhibin Li, and Yinhai Wang. A hierarchical model for taxis' customer searching behaviors using gps trajectory data. In *Proceedings of 95th Annual Meeting of the Transportation Research Board*, January 2016.

[38] M. Keith Chen. Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 455–455, New York, NY, USA, 2016. ACM.

[39] Suiming Guo, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. Understanding ride-on-demand service: Demand and dynamic pricing. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 509–514, March 2017.

[40] Le Chen, Alan Mislove, and Christo Wilson. Peeking beneath the hood of uber. In *Proceedings of the 2015 Internet Measurement Conference*, IMC '15, pages 495–508, New York, NY, USA, 2015. ACM.

[41] Yujie Zhang, Hongwei Guo, Cheng Li, Wuhong Wang, Xiaobei Jiang, and Yang Liu. Which one is more attractive to traveler, taxi or tailored taxi? an empirical study in china. *Procedia Engineering*, 137(Supplement C):867 – 875, 2016. Green Intelligent Transportation System and Safety.

[42] Joseph P. Schwieterman and Matthew Michel. Have app will travel: comparing price & speed of fifty cta & uberpool trips in chicago. Technical report, https://las.depaul.edu/centers-and-institutes/chaddick-institute-for-metropolitan-development/research-and-publications/Documents/Have2016.

[43] Felipe F. Dias, Patrícia S. Lavieri, Venu M. Garikapati, Sebastian Astroza, Ram M. Pendyala, and Chandra R. Bhat. A behavioral choice model of the use of car-sharing and ride-sourcing services. *Transportation*, 44(6):1307–1323, Nov 2017.

[44] Reuben Fischer-Baum and Carl Bialik. Uber is taking millions of manhattan rides away from taxis. Technical report, https://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/, October 2015.

[45] Carl Bialik, Andrew Flowers, Reuben Fisher-Baum, and Dhrumil Metha. Uber is serving new york's outer boroughs more than taxis are. Technical report, https://fivethirtyeight.com/features/uber-is-serving-new-yorks-outer-boroughs-more-than-taxis-are/, August 2015.

[46] Judd Cramer and Alan B. Krueger. Disruptive change in the taxi business: The case of uber. *American Economic Review*, 106(5):177–82, May 2016.

[47] Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. Working Paper 22627, National Bureau of Economic Research, September 2016.

[48] Xiqun (Michael) Chen, Majid Zahiri, and Shuaichao Zhang. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies*, 76(Supplement C):51 – 70, 2017.

[49] Anastasios Noulas, Vsevolod Salnikov, Desislava Hristova, Cecilia Mascolo, and Renaud Lambiotte. Developing and deploying a taxi price comparison mobile app in the wild: Insights and challenges. *CoRR*, abs/1701.04208, 2017.

[50] Ryan Hughes and Don MacKenzie. Transportation network company wait times in greater seattle, and relationship to socioeconomic indicators. *Journal of Transport Geography*, 56(Supplement C):36 – 44, 2016.

[51] Suiming Guo, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. Understanding passenger reaction to dynamic prices in ride-on-demand service. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 42–45, March 2017.

[52] Josep Maria Salanova, Miquel Estrada, Georgia Aifadopoulou, and Evangelos Mitsakis. A review of the modeling of taxi services. *Procedia - Social and Behavioral Sciences*, 20(Supplement C):150 – 161, 2011. The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transport.

[53] Josep Maria Salanova, Miquel Estrada Romeu, and Carles Amat. Aggregated modeling of urban taxi services. *Procedia - Social and Behavioral Sciences*, 160(Supplement C):352 – 361, 2014. XI Congreso de Ingenieria del Transporte (CIT 2014).

[54] Nelson D. Chan and Susan A. Shaheen. Ridesharing in north america: Past, present, and future. *Transport Reviews*, 32(1):93–112, 2012.

[55] Masabumi Furuhata, Maged Dessouky, Fernando Ordóñez, Marc-Etienne Brunet, Xiaoqing Wang, and Sven Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57(Supplement C):28 – 46, 2013.

[56] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, and F. Herrera. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5(1):146 – 166, 2011.

[57] Bruce Schaller. Elasticities for taxicab fares and service availability. *Transportation*, 26(3):283–297, Aug 1999.

[58] Hai Yang and Teng Yang. Equilibrium properties of taxi markets with search frictions. *Transportation Research Part B: Methodological*, 45(4):696 – 713, 2011.

[59] Camille Kamga, M. Anil Yazici, and Abhishek Singhal. Analysis of taxi demand and supply in new york city: implications of recent taxi regulations. *Transportation Planning and Technology*, 38(6):601–625, 2015.

[60] Annick Lacombe and Catherine Morency. Modeling taxi trip generation using gps data: the montreal case. In *Proceedings of 95th Annual Meeting of the Transportation Research Board*, January 2016.

[61] Jun Guan Neoh, Maxwell Chipulu, and Alasdair Marshall. What encourages people to carpool? an evaluation of factors with meta-analysis. *Transportation*, 44(2):423–447, Mar 2017.

[62] Abel Brodeur and Kerry Nield. An empirical analysis of taxi, lyft and uber rides: Evidence from weather shocks in nyc. *Journal of Economic Behavior & Organization*, 152:1 – 16, 2018.

[63] Wenbo Zhang, Tho V. Le, Satish V. Ukkusuri, and Ruimin Li. Influencing factors and heterogeneity in ridership of traditional and app-based taxi systems. *Transportation*, Oct 2018.

[64] Luis Moreira-Matias, Joao Gama, Michel Ferreira, Joao Mendes-Moreira, and Luis Damas. Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1393–1402, Sept 2013.

[65] K. Zhao, D. Khryashchev, J. Freire, C. Silva, and H. Vo. Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 833–842, Dec 2016.

[66] Xinwu Qian, Satish V Ukkusuri, Chao Yang, and Fenfan Yan. A model for short-term taxi demand forecasting accounting for spatio-temporal correlations. In *Transportation Research Board 96th Annual Meeting*, 2017.

[67] Xinwu Qian, Satish V Ukkusuri, Chao Yang, and Fenfan Yan. Forecasting short-term taxi demand using boosting-gcrf. In *6th ACM SIGKDD International Workshop on Urban Computing (UrbComp 2017)*, 2017.

[68] Siyu Liao, Liutong Zhou, Xuan Di, Bo Yuan, and Jinjun Xiong. Large-scale short-term urban taxi demand forecasting using deep learning. In *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*, ASPDAC '18, pages 428–433, Piscataway, NJ, USA, 2018. IEEE Press.

[69] Sabiheh Sadat Faghih, Abolfazl Safikhani, Bahman Moghimi, and Camille Kamga. Predicting short-term demand of uber using spatio-temporal modeling, case study: New york city. In *Transportation Research Board 97th Annual Meeting*, 2018.

[70] A. Afian, A. Odoni, and D. Rus. Inferring unmet demand from taxi probe data. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 861–868, Sept 2015.

[71] Hai Yang, S.C. Wong, and K.I. Wong. Demand–supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research Part B: Methodological*, 36(9):799 – 819, 2002.

[72] K.I. Wong, S.C. Wong, and Hai Yang. Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research Part B: Methodological*, 35(9):819 – 842, 2001.

[73] Josep Maria Salanova Grau and Miquel Estrada. Social optimal shifts and fares for the barcelona taxi sector. *Transport Policy*, 2017.

[74] Patrícia S. Lavieri, Felipe F. Dias, Natalia Ruiz Juri, James Kuhr, and Chandra R. Bhat. A model of ridesourcing demand generation and distribution. *Transportation Research Record*, 0(0):0361198118756628, 2018.

[75] Yang Yingjun, Hua Cui, Zhang Shaoyang, and Yang Yingjun. A prediction model of the number of taxicabs based on wavelet neural network. *Procedia Environmental Sciences*, 12(Part B):1010 – 1016, 2012. 2011 International Conference of Environmental Science and Engineering.

[76] R. Gelda, K. Jagannathan, and G. Raina. Taxi dispatches using supply forecasting: A time-series based approach. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 1333–1340, Dec 2016.

[77] Daniel Flores-Guri. An economic analysis of regulated taxicab markets. *Review of Industrial Organization*, 23(3):255–266, Dec 2003.

[78] Henry S. Farber. Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of Political Economy*, 113(1):46–82, 2005.

[79] Henry S. Farber. Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *American Economic Review*, 98(3):1069–82, June 2008.

[80] Sumit Agarwal (department Of Finance. Labor supply decisions of singaporean cab drivers, 2013.

[81] Vincent Martin. When to quit: Narrow bracketing and reference dependence in taxi drivers. *Journal of Economic Behavior & Organization*, 144:166 – 187, 2017.

[82] K. Nar, L. J. Ratliff, and S. Sastry. Learning prospect theory value function and reference point of a sequential decision maker. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5770–5775, Dec 2017.

[83] John Joseph Horton. Labor market equilibration : Evidence from uber, 2017.

[84] W.Y. Szeto, R.C.P. Wong, and W.H. Yang. Guiding vacant taxi drivers to demand locations by taxi-calling signals: A sequential binary logistic regression modeling approach and policy implications. *Transport Policy*, 2018.

[85] R. M. N. T. Sirisoma, Siu C. Wong, William H. K. Lam, Donggen Wang, Hai Yang, and Peng Zhang. Empirical evidence for taxi customer-search model. *Proceedings of the Institution of Civil Engineers - Transport*, 163(4):203–210, 2010.

[86] R.C.P. Wong, W.Y. Szeto, S.C. Wong, and Hai Yang. Modelling multi-period customer-searching behaviour of taxi drivers. *Transportmetrica B: Transport Dynamics*, 2(1):40–59, 2014.

[87] Zhong Zheng, Soora Rasouli, and Harry Timmermans. Modeling taxi driver anticipatory behavior. *Computers, Environment and Urban Systems*, 69:133 – 141, 2018.

[88] Ryan C.P. WONG, W.Y. SZETO, and S.C. WONG. Sequential logit approach to modeling the customer-search decisions of taxi drivers. *Asian Transport Studies*, 3(4):398–415, 2015.

[89] R.C.P. Wong, W.Y. Szeto, and S.C. Wong. A cell-based logit-opportunity taxi customer-search model. *Transportation Research Part C: Emerging Technologies*, 48(Supplement C):84 – 96, 2014.

[90] R.C.P. Wong, W.Y. Szeto, and S.C. Wong. A two-stage approach to modeling vacant taxi movements. *Transportation Research Part C: Emerging Technologies*, 59(Supplement C):147 – 163, 2015. Special Issue on International Symposium on Transportation and Traffic Theory.

[91] Jiancheng Long, W.Y. Szeto, Jie Du, and R.C.P. Wong. A dynamic taxi traffic assignment model: A two-level continuum transportation system approach. *Transportation Research Part B: Methodological*, 100(Supplement C):222 – 254, 2017.

[92] Yingjie Zhang, Beibei Li, Ramayya Krishnan, and Siyuan Liu. Learning from the offline trace: A case study of the taxi industry. In *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*, United States, 2015. Association for Information Systems.

[93] Hai Yang, C.S. Fung, K.I. Wong, and S.C. Wong. Nonlinear pricing of taxi services. *Transportation Research Part A: Policy and Practice*, 44(5):337 – 348, 2010.

[94] Mariano Gallo. Improving equity of urban transit systems with the adoption of origin-destination based taxi fares. *Socio-Economic Planning Sciences*, 2017.

[95] Cheng Zeng and Nir Oren. Dynamic taxi pricing. pages 1135–1136, 8 2014. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License. This research is supported by the award made by the RCUK Digital Economy theme to the dot.rural Digital Economy Hub, award reference: EP/G066051/1.; European Conference on Artificial Intelligence (ECAI-2014) ; Conference date: 18-08-2014 Through 22-08-2014.

[96] Xinwu Qian and Satish V. Ukkusuri. Time-of-day pricing in taxi markets. *IEEE Transactions on Intelligent Transportation Systems*, 18(6):1610–1622, June 2017.

[97] Steven Suranovic. Surge Pricing and Price Gouging: Public Misunderstanding as a Market Imperfection. Working Papers 2015-20, The George Washington University, Institute for International Economic Policy, May 2016.

[98] Jonathan Hall, Cory Kendrick, and Chris Nosko. The effects of uber's surge pricing: A case study. Technical report, Uber Inc., http://economicsforlife.ca/wp-content/uploads/2015/10/effects_of_ubers_surge_pricing.pdf, 2015.

[99] S. Guo, C. Chen, Y. Liu, K. Xu, and D.M. Chiu. It can be cheaper: Using price prediction to obtain better prices from dynamic pricing in ride-on-demand services. In *ACM International Conference Proceeding Series*, pages 146–155. Association for Computing Machinery, 2017.

[100] W. Zhang, D. Kumar, and S. V. Ukkusuri. Exploring the dynamics of surge pricing in mobility-on-demand taxi services. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1375–1380, Dec 2017.

[101] Carlos F. Daganzo. An approximate analytic model of many-to-many demand responsive transportation systems. *Transportation Research*, 12(5):325 – 333, 1978.

[102] Ying Shi and Zhaotong Lian. Equilibrium strategies and optimal control for a double-ended queue. *Asia-Pacific Journal of Operational Research*, 33(03):1650022, 2016.

[103] A. Anwar, M. Volkov, and D. Rus. Changinow: A mobile application for efficient taxi allocation at airports. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 694–701, Oct 2013.

[104] G. Qi, G. Pan, S. Li, Z. Wu, D. Zhang, L. Sun, and L. T. Yang. How long a passenger waits for a vacant taxi – large-scale taxi trace mining for smart cities. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pages 1029–1036, Aug 2013.

[105] C.F. Manski and J.D. Wright. The nature of equilibrium in the market for taxi services. *Transportation Research Record*, 619:11–15, 1976.

[106] Hai Yang, Min Ye, Wilson H. Tang, and S.C. Wong. Regulating taxi services in the presence of congestion externality. *Transportation Research Part A: Policy and Practice*, 39(1):17 – 40, 2005.

[107] Liteng Zha, Yafeng Yin, and Yuchuan Du. Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Procedia*, 23(Supplement C):2 – 21, 2017. Papers Selected for the 22nd International Symposium on Transportation and Traffic Theory Chicago, Illinois, USA, 24-26 July, 2017.

[108] Josep Maria Salanova Grau and Miquel Angel Estrada Romeu. Agent based modelling for simulating taxi services. *Procedia Computer Science*, 52(Supplement C):902 – 907, 2015. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).

[109] David K. George and Cathy H. Xia. Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research*, 211(1):198 – 207, 2011.

[110] Siddhartha Banerjee, Daniel Freund, and Thodoris Lykouris. Pricing and optimization in shared vehicle systems: An approximation framework. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 517–517, New York, NY, USA, 2017. ACM.

[111] Rick Zhang and Marco Pavone. Control of robotic mobility-on-demand systems: A queueing-theoretical perspective. *The International Journal of Robotics Research*, 35(1-3):186–203, 2016.

[112] Ying Shi and Zhaotong Lian. Optimization and strategic behavior in a passenger–taxi service system. *European Journal of Operational Research*, 249(3):1024 – 1032, 2016.

[113] Ying Shi, Z. Lian, and Weixin Shang. Study of a passenger-taxi queueing system with nonzero matching time. In *2015 12th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–5, June 2015.

[114] Mohsen Ramezani and Mehdi Nourinejad. Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach. *Transportation Research Part C: Emerging Technologies*, 2017.

[115] S K Jason Chang, Chi-Hsuan Wu, and Chih-Hsu Lin. Optimization of fleet size for gps taxi. In *Proceedings of 91st Annual Meeting of the Transportation Research Board*, January 2012.

[116] Wenbo Zhang and Satish V. Ukkusuri. Optimal fleet size and fare setting in emerging taxi markets with stochastic demand. *Computer-Aided Civil and Infrastructure Engineering*, 31(9):647–660, 2016.

[117] Niels Agatz, Alan Erera, Martin Savelsbergh, and Xing Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295 – 303, 2012.

[118] Douglas O. Santos and Eduardo C. Xavier. Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2885–2891. AAAI Press, 2013.

[119] Douglas O. Santos and Eduardo C. Xavier. Taxi and ride sharing: A dynamic dial-a-ride problem with money as an incentive. *Expert Systems with Applications*, 42(19):6728 – 6737, 2015.

[120] Ali Shemshadi, Quan Z. Sheng, and Wei Emma Zhang. *A Decremental Search Approach for Large Scale Dynamic Ridesharing*, pages 202–217. Springer International Publishing, Cham, 2014.

[121] Pedro M. d'Orey, Ricardo Fernandes, and Michel Ferreira. Empirical evaluation of a dynamic and distributed taxi-sharing system. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 140–146, Sept 2012.

[122] Erhun Ozkan and Amy R Ward. Dynamic matching for real-time ridesharing. *Available at SSRN*, June 2017.

[123] Niels A.H. Agatz, Alan L. Erera, Martin W.P. Savelsbergh, and Xing Wang. Dynamic ride-sharing: A simulation study in metro atlanta. *Transportation Research Part B: Methodological*, 45(9):1450 – 1464, 2011. Select Papers from the 19th ISTTT.

[124] Alexander J. Schindele. A ridesharing analysis with a hitchhiking modification applied to taxi trips in new york.

[125] Chi-Chung Tao. Dynamic taxi-sharing service using intelligent transportation system technologies. In *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, pages 3209–3212, Sept 2007.

[126] Chi-Chung Tao and Chun-Ying Chen. Heuristic algorithms for the dynamic taxipooling problem based on intelligent transportation system technologies. In *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, volume 3, pages 590–595, Aug 2007.

[127] Ker-Tsung Lee, Da-Jie Lin, and Pei-Ju Wu. Planning and design of a taxipooling dispatching system. *Transportation Research Record: Journal of the Transportation Research Board*, 1903:86–95, 2005.

[128] Shuo Ma, Yu Zheng, and Ouri Wolfson. T-share: A large-scale dynamic taxi ridesharing service. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, pages 410–421, April 2013.

[129] Shuo Ma and Ouri Wolfson. Analysis and evaluation of the slugging form of ridesharing. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, pages 64–73, New York, NY, USA, 2013. ACM.

[130] Mitja Stiglic, Niels Agatz, Martin Savelsbergh, and Mirko Gradisar. The benefits of meeting points in ride-sharing systems. *Transportation Research Part B: Methodological*, 82(Supplement C):36 – 53, 2015.

[131] Roger Lloret-Batlle, Neda Masoud, and Daisik Nam. Peer-to-peer ridesharing with ride-back on high-occupancy-vehicle lanes. *Transportation Research Record: Journal of the Transportation Research Board*, 2668:21–28, 2017.

[132] Baoxiang Li, Dmitry Krushinsky, Hajo A. Reijers, and Tom Van Woensel. The share-a-ride problem: People and parcels sharing taxis. *European Journal of Operational Research*, 238(1):31 – 40, 2014.

[133] Xiaoshu Ma, Jinming Ding, Wei Wang, Xuepeng Hua, and Yujia Peng. A taxipooling system with equity consideration. In *PROCEEDINGS OF The 14th International Conference on Computers in Urban Planning and Urban Management*, 2015.

[134] Ouri Wolfson and Jane Lin. Fairness versus optimality in ridesharing. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, pages 118–123, May 2017.

[135] Po-Yu Chen, Je-Wei Liu, and Wen-Tsuen Chen. A fuel-saving and pollution-reducing dynamic taxi-sharing protocol in vanets. In *2010 IEEE 72nd Vehicular Technology Conference - Fall*, pages 1–5, Sept 2010.

[136] Biying Yu, Ye Ma, Meimei Xue, Baojun Tang, Bin Wang, Jinyue Yan, and Yi-Ming Wei. Environmental benefits from ridesharing: A case of beijing. *Applied Energy*, 191(Supplement C):141 – 152, 2017.

[137] Huayu Xu, Jong-Shi Pang, Fernando Ordóñez, and Maged Dessouky. Complementarity models for traffic equilibrium with ridesharing. *Transportation Research Part B: Methodological*, 81(Part 1):161 – 182, 2015.

[138] Jamal Yousaf, Juanzi Li, Lu Chen, Jie Tang, Xiaowen Dai, and John Du. *Ride-Sharing: A Multi Source-Destination Path Planning Approach*, pages 815–826. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[139] Jamal Yousaf, Juanzi Li, Lu Chen, Jie Tang, and Xiaowen Dai. Generalized multipath planning model for ride-sharing systems. *Frontiers of Computer Science*, 8(1):100–118, Feb 2014.

[140] Qing Ye, Changxi Ma, Ruichun He, Qiang Xiao, and Wei Zhang. Multi-objective optimisation for taxi ridesharing route based on non-dominated sorting genetic algorithm. *International Journal of Wireless and Mobile Computing*, 8(3):262–270, 2015.

[141] Mehdi Nourinejad and Matthew J. Roorda. Agent based model for dynamic ridesharing. *Transportation Research Part C: Emerging Technologies*, 64(Supplement C):117 – 132, 2016.

[142] Luis M. Martinez, Gonçalo H. A. Correia, and José M. Viegas. An agent-based simulation model to assess the impacts of introducing a shared-taxi system: an application to lisbon (portugal). *Journal of Advanced Transportation*, 49(3):475–495, 2015. ATR-14-0036.R1.

[143] Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H. Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294, 2014.

[144] Tal Altshuler, Rachel Katoshevski, and Yoram Shiftan. Ride sharing and dynamic networks analysis. *CoRR*, abs/1706.00581, 2017.

[145] Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467, 2017.

[146] Xinwu Qian, Wenbo Zhang, Satish V. Ukkusuri, and Chao Yang. Optimal assignment and incentive design in the taxi group ride problem. *Transportation Research Part B: Methodological*, 103(Supplement C):208 – 226, 2017. Green Urban Transportation.

[147] ALEX ROSENBLAT. Uber's Phantom Cabs, accessed June, 2017. Available online at motherboard.vice.com/en_us/article/mgbz5a/ubers-phantom-cabs.

[148] The Guardian. Uber denies misleading passengers with 'phantom cars' on app, accessed June, 2017. Available online at https://www.theguardian.com/technology/2015/jul/30/uber-denies-misleading-passengers-with-phantom-cars-on-app.

[149] CONSTANCE L. WOOD and MICHELE M. ALTAVELA. Large-sample results for kolmogorov-smirnov statistics for discrete distributions. *Biometrika*, 65(1):235–239, 1978.

[150] Michael Theis. The uber effect: Austin taxi rides drop dramatically in past year, 2016. Available online at https://www.bizjournals.com/austin/news/2016/01/19/the-Uber-effect-austin-taxi-rides-drop.html.

[151] Jack Newsham and Dan Adams. Amid fight with uber, lyft, boston taxi ridership plummets, 2015. Available online at https://www.bostonglobe.com/business/2015/08/19/boston-taxi-ridership-down-percent-this-year/S9dZMELMye6puzTTYoDIrL/story.html.

[152] Paul Bradley Carr. As NY floods, "Robin Hood" Uber robs from the rich and... Nope, that's about it, accessed August, 2017. Available online at `https://pando.com/2012/10/31/assholes-shrug/`.

[153] Nick Bilton. Customers Out in the Cold Balk at Uber Surge Pricing, accessed August, 2017. Available online at `https://bits.blogs.nytimes.com/2013/12/16/customers-out-in-the-cold-balk-at-uber-surge-pricing/?_r=0`.

[154] Jay Hathaway. Uber Turned on Surge Pricing for People Fleeing Sydney Hostage Scene, accessed August, 2017. Available online at http://gawker.com/uber-turned-on-surge-pricing-for-people-fleeing-sydney-1671193132.

[155] Charles Riley. Uber criticized for surge pricing after London terror attack, accessed August, 2017. Available online at http://money.cnn.com/2017/06/04/technology/uber-london-attack-surge-pricing/index.html.

[156] Mike Isaac. Uber Reaches Deal With New York on Surge Pricing in Emergencies, accessed August, 2017. Available online at https://bits.blogs.nytimes.com/2014/07/08/uber-reaches-agreement-with-n-y-on-surge-pricing-during-emergencies/?src=twr.

[157] Uber Newsroom. Partnership with american red cross to support cities and citizens during disasters, accessed August, 2017. Available online at https://newsroom.uber.com/partnership-with-american-red-cross-to-support-cities-and-citizens-during-disasters/.

[158] Uber Help. What is dynamic pricing?, accessed August, 2017. Available online at https://help.uber.com/h/34212e8b-d69a-4d8a-a923-095d3075b487.

[159] Abhay Valiyaveettil, Afroz Mohammad, Amy Zhang, Anisha Kaul, Brahmi Kavuri, and Dan Xie. Uber dynamic pricing: A research project on the reasons behind uber dynamic pricing and the underlying technologies., accessed August, 2017. Available online at https://www.andrew.cmu.edu/user/avaliyav/Uber_new/.

[160] Bill Gurley. A deeper look at uber's dynamic pricing mode, accessed August, 2017. Available online at https://newsroom.uber.com/guest-post-a-deeper-look-at-ubers-dynamic-pricing-model/.

[161] Gregory Bresiger. Investors see opportunity as taxi medallion prices 'bottom out', Oct 2017.

[162] Hua Cai, Xiaowei Zhan, Ji Zhu, Xiaoping Jia, Anthony S.F. Chiu, and Ming Xu. Understanding taxi travel patterns. *Physica A: Statistical Mechanics and its Applications*, 457(Supplement C):590 – 597, 2016.

[163] M. M. Vazifeh, P. Santi, G. Resta, S. H. Strogatz, and C. Ratti. Addressing the minimum fleet problem in on-demand urban mobility. *Nature*, 557(7706):534–538, 2018.

[164] Liteng Zha, Yafeng Yin, and Hai Yang. Economic analysis of ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, 71(Supplement C):249 – 266, 2016.

[165] T. Heilker and G. Sieg. A duopoly of transportation network companies and traditional radio-taxi dispatch service agencies. *European Journal of Transport and Infrastructure Research*, 18(2):196–212, 2018. cited By 0.

[166] J. Michael Harrison. Assembly-like queues. *Journal of Applied Probability*, 10(2):354–367, 1973.

[167] Nikolas Geroliminis and Carlos F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759 – 770, 2008.

[168] David R. Alexander, I. M. Premachandra, and Toshikazu Kimura. Transient and asymptotic behavior of synchronization processes in assembly-like queues. *Annals of Operations Research*, 181(1):641–659, Dec 2010.

[169] Balaji Prabhakar, Nicholas Bambos, and T. S. Mountford. The synchronization of poisson processes and queueing networks with service and synchronization nodes. *Advances in Applied Probability*, 32(3):824–843, 2000.

[170] F.P. Kelly. *Reversibility and Stochastic Networks.*

[171] Lawrence D. Brown and Linda H. Zhao. A test for the poisson distribution. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 64(3):611–625, 2002.

VITA

VITA

Wenbo Zhang is a motivated researcher with a broad knowledge of mathematical programming, algorithm design, behavioral modeling, and data mining techniques. In his PhD, he developed significant contributions in the area of urban taxi system, varying from big data acquisition, behavioral modeling, mobility patterns, to stochastic system modeling. He has several publications in peer-reviewed top journals and conferences in the fields of not only Transportation but also Big Data. He is a recipient of the China Scholarship Council fellowship (2013 to 2017), ThinkSwiss Research Scholarship (2017), Purdue Engineering Graduate Students Travel Grant (2017), and Erin Flanagan Travel Grant (2018). Moreover, he has experience in several transportation research projects funded by National Science Foundation (NSF) and Indiana Department of Transportation (INDOT), related to social media data collection, traffic state estimation during hurricane, and traffic safety analyses for right-turn lane designs. Before joining Urban Mobility and Network Intelligence (UMNI) Lab directed by Dr. Satish V. Ukkusuri at Purdue University, I competed my Master's degree in transportation planning and management at Southeast University in 2013 and my Bachelor's degree in Transportation at China University of Mining and Technology in 2010.