Search Frictions, Network Effects and Spatial Competition: Taxis versus Uber

Bo Bian*
Pennsylvania State University

April 15, 2018

Abstract

In this paper, I model the search and matching process among passengers, taxis and Uber drivers in New York City to analyze the matching efficiency taking into account network effects and supply competition. Drivers make dynamic spatial search decisions to supply rides across locations (platforms) and passengers make static discrete choice decisions among taxi and Uber. Network effects occur if increased participation of one side impacts searches of the other side in the same location. I model network effects by adding demand and supply to both sides' decisions. I use the nonstationary oblivious equilibrium to estimate the dynamic model and analyze frictions as mismatches between drivers and passengers. I find significant network effects. Then, I show that network effects and supply competition, in addition to fixed pricing structure (of taxi), have extensive effects on frictions and welfare in three counterfactual scenarios. The first eliminates the Uber surge multiplier, the second improves traffic conditions and the third decreases Uber drivers by 30 percent. I find that without surge pricing, Uber's mismatches increase by 3,152 in a day shift. Improving traffic condition reduces matching frictions for both taxis and Uber. It increases searches of drivers and increases taxis' pickups by 10.6 percent. Taxis' profit increases by $ 231,100 and Uber's increases by $5,110. Restricting Uber's supply increases frictions for both taxis and Uber due to less competition. But it increases taxis' pickups by 1,660 ($ 34,300) compared to 7,004 ($ 119,320) loss of Uber's trips. Consumers are worse off after the regulatory policy but better off in the first two scenarios. Most importantly, with or without accounting for network effects generates different simulation results and could lead to opposite conclusions.

Key Words: spatial competition, network effects, search and matching, regulation. (*JEL: C73, D83, L90, R12*)

*Department of Economics, The Pennsylvania State University. Email: **bub186@psu.edu**. I am very grateful to my advisor Mark Roberts for his constant support and valuable comments. I would also like to thank Peter Newberry, Charles Murry, Joris Pinkse, Daniel Grodzicki and participants in PSU brownbag seminar for their comments and discussions. I am solely responsible for any errors.

1 Introduction

Frictions play an important role in explaining the failure of market clearing. When sellers and buyers meet and trade with each other, information imperfections about potential trading partners, heterogeneities, slow mobility and congestion from large numbers can prevent some potential traders on one side of the market from contacting potential traders on the other side, leaving some buyers and sellers unable to trade. The early search and matching literature use a reduced form matching function to capture effects of frictions on equilibrium outcomes of bilateral trades[1]. The source of frictions underling such a function is not explicitly modelled.

In more recent literature, the microfoundation of the matching function is studied. Lagos (2000) builds a model of taxis' spatial search for passengers and finds that even without imperfection information and random search assumption, aggregate mismatches over locations arise endogenously as outcomes of drivers' optimal search decisions. Specifically, when one location is more profitable than another, taxis may overcrowd that location leaving another location with unserved passengers. Similar to the idea in Lagos (2000), Buchholz (2016) empirically studies search frictions in the taxi industry as consequence of price regulation which fails to coordinate cross-market demand and supply leaving empty taxis in some areas and excess demand in other areas. However, both Lagos (2000) and Buchholz (2016) emphasize the prices as a source of spacial mismatches and mainly focus on the supply side of the matching process.

This paper studies the matching efficiency in the taxi and ridesharing industry [2]. I contribute to previous literature by developing a richer demand side accounting for network effects in the way that decisions of both drivers and passengers depend on the size of the other side of the market. Treating a location as a platform, such feedback loop is called indirect network effect (INE) in the literature[3]. I study how this non-price factor influences matching efficiency of the market. In details, on the supply side, drivers make spatial search decisions among locations by choosing the one

[1] Blanchard and Diamond (1989), Pissarides (1990), and Mortensen and Pissarides (1994) are examples.

[2] Taxi industry is well known for existence of matching frictions such that some areas have excess demand whereas some have excess supply. This industry is ideal for analyzing search and matching frictions for several reasons. First, search decisions are made by decentralized individuals without coordination. Second, the taxi market is highly regulated in fares and medallions. Third, there is no preference heterogeneity among drivers and passengers.

[3] I refer to cross sides externality as indirect network effect (INE) and the same side externality as direct network effect (DNE) for the rest of this paper.

with the highest expected profit taking into account both demand and supply in the location. The demand positively affects expected profit by increasing matching probability, whereas the supply competition negatively affects expected profit by decreasing matching probability. On the demand side, passengers make a static discrete choice decision among taxi, Uber and an outside option. From the passengers' perspective, the supply of cars changes the matching probabilities and waiting time which affects their utility and choice decisions. This interdependence between demand and supply, if exists, has effects on matching efficiencies. For instance, if the INE is positive, an increase of supply in a location with excess supply will increase demand which further attracts more drivers. Supposing the increased demand is less (more) than increased supply, excess supply will be larger (smaller) at new equilibrium than without network effects. In other words, network effects provide an extra channel which may exacerbate or alleviate matching frictions caused by inefficient prices.

Instead of modelling only taxi drivers' search decisions as in Buchholz (2016), I further allow the model to include ridesharing company Uber. Taxis and Uber compete for passengers in each location by providing products with different prices and qualities (i.e. waiting time, matching probability). I also allow the matching technology within location to differ between taxis and Uber. The benefits of modelling Uber are twofold. First, in addition to the DNE within taxis that matching probability decreases in number of taxi drivers, Uber drivers affect taxis' profits in a different way by affecting the demand for taxis. Second, with the rapid growth of ridesharing industry, there are many real world debates about the influence of Uber on traditional taxis' profits and its regulation. My duopoly model can simulate several policy outcomes and makes suggestions to the policy makers.

To empirically analyze the model, I use data on trip records of taxis and Uber from the New York City Taxi and Limousine Commission(TLC) and collected Uber's surge multiplier in April 2016. This dataset provides detailed information on taxis' trips including pickup/dropoff locations and time for trips and brief information on Uber's trips including only pickup areas and time. This data is very limited since it does not provide information on supply and demand levels in the matching process. I borrows the strategy from Buchholz (2016) to estimate equilibrium demand and supply out of a dynamic spatial search model for both taxi and Uber drivers. Due to the large number of drivers in this dynamic game, we apply the concept of nonstationary oblivious equilibrium proposed by Weintraub, Benkard and Jeziorski (2008) to solve

the equilibrium. The idea of OE is that, instead of competing with each other, taxi and Uber drivers are atomistic and compete against deterministic paths of distribution of other drivers in equilibrium.

Finally, as implication of my model estimates, I simulate several counterfactuals for the interests of both matching efficiency of the market and policy issues in the real world. During the sample period of April 2016, Uber is growing rapidly and the total Uber licensed drivers outnumber the taxis[4]. There are mainly two complaints about the growth of Uber. First, people blame that it contributes to traffic congestion. Second, it causes the taxis' profits to drop[5]. In 2015, the city governor proposed to solve these problems by capping the growth of Uber though not implemented. Now, city governor thinks of regulating Uber again and they also propose to charge congestion pricing to improve traffic in Manhattan[6].

Thus, I focus three main counterfactuals. In the first counterfactual, I eliminate Uber's surge multiplier to study to what extent the flexible pricing improves spatial matching efficiency. In the second one, I improves the traffic condition to study the magnitude of traffic condition's impact on matching efficiency. In the third one, I simulate the regulatory policy of restricting Uber's supply and analyze how competitiveness affects matching efficiency and profit changes of taxis. In each simulation, I predict and compare market outcomes with and without network effect. But, I do not endogenously model traffic conditions so that the simulation results are partial effects.

I find that without surge multiplier, Uber's cross location mismatches increase whereas taxis' mismatches decrease. Lower price of Uber make the market more competitive and competition improves taxi's matching efficiency. After improving traffic condition, I find that matching efficiencies increase for both firms. Pickups of taxis increase by 10.6% whereas Uber's pickups decrease by 2.21%. Though I do not endogenously model traffic condition in this paper, the second simulation suggests importance of traffic speeds in spatial matching efficiency. In the last counterfactual, after restricting Uber's supply, taxis' demand and pickups increase slightly in comparison to decline of Uber's pickups. The cross location mismatches of taxis increase a bit.

[4]There are $26,000$ Uber's licensed vehicles in comparison to $13,000$ taxi medallions. Uber completes 4.6 million trips and taxis complete 13 million in the sample month

[5]For example, the auction price of an independent unrestricted medallion dropped from \$ 0.7 million in 2011 to \$0.5 million in 2016.

[6]https://www.timeout.com/newyork/news/the-city-council-finally-remembered-that-uber-needs-to-be-regulated-in-nyc-030218

All results indicate the importance of network effects on the simulation outcomes and policy conclusions.

The rest of this paper is organized as follows. Section 2 discusses prior literature in detail. In section 3, I present a simplified model demonstrating why the network effects matter for matching efficiency. Section 4 and 5 introduce industry background and the data separately. Section 6 presents the empirical model. Estimation strategy is in section 7 and the results are in section 8. In section 9, I simulate and discuss counterfactuals. Section 10 concludes this paper.

2 Literature

This paper is built upon two streams of literature, network effects and search and matching. This paper contributes to the network effects literature by modelling both DNE and INE. DNE measure the externalities of other agents from the same side of market on agent's decision. For example, other drivers in the same location would decrease the chance of being matched. INE measures the impact of agents from the other side of market on agent's decision. For example, high demand increases supply of drivers and more drivers will increase passengers' choice probability of taxi(or Uber). If INE is bilateral, it forms a feedback loop for evolution of both sides. The theoretical literature of network effects begins with Katz and Shapiro(1985) and follows by Farrell and Saloner (1986), Chou and Shy (1990), Church and Gandal (1992), Rochet and Tirole(2003, 2006) and Amstrong (2006).

The empirical works on network effects begin with Gandal (1994), Saloner and Shepard (1995) and grow rapidly in recent years. For instance, Gandal, Kende, and Rob (2000) develop a dynamic model of consumer's adoption of CD player and software entry to estimate the feedback in CD industry. Ohashi (2003) and Park (2004) study network effects in the U.S. home VCR market. Nair, Chintagunta, and Dube(2004) quantify the network effects in the PDA market. Clements and Ohashi (2005), Corts and Lederman (2009) focus on network effects in the video game industry. Rysman (2004) estimates the network effects in Yellow Pages market and how it is related to market concentration and antitrust policy. Ackerberg and Gowrisankaran (2006) estimate the importance of network effect in the ACH banking industry. Dubé, Hitsch, and Chintagunta (2010) study network effects in video game market and its tipping effects. Lee (2013) also studies game industry but focus on software exclusivity. Luo (2016)

studies network effects in smartphone industry and how it affects carriers' dynamic penetration pricing strategy.

Most of these empirical works focus on indirect network effect of a two-sided platform and ignore the direct network effects. Or they use network size of one side to estimate joint effects of INE and DNE. Goolsbee and Klenow (2002) is one paper focusing on only direct network effect in the diffusion of home computers. Chu and Manchanda (2016) is one recent paper trying to estimates and distinguish both direct and indirect network effects in e-commerce platform (Alibaba). My paper contributes to the literature by quantifying both direct and indirect network effects. I allow the sizes of agents from both sides of the market to affect decisions of individuals on either side.

This paper also contributes to the search and matching literature by adding network effects to the model. Early search and matching model use reduced form matching function to introduce frictions that prevent the market from clearing (Blanchard and Diamond(1989), Pissarides(1984), Mortensen and Pissarides (1999)). Microfoundations of the matching function are introduced, for example, as coordination failures in ball-and-urn problem (Butters (1977) and Burdett, Shi and Wright (2001)). Lagos (2000) develops a spacial search model of taxis without imperfect information and random search assumptions showing that frictions arise in the aggregate matching function endogenously as outcomes of drivers' search decisions. Specifically, when prices are fixed and one location is more lucrative than other locations, drivers will overcrowd that location leaving other locations undersupplied. Coexistence of excess demand and excess supply reflect frictions in the aggregate matching function. Buchholz (2016) extends Lagos (2000) and builds an empirical model with non-stationary drivers' dynamics and price-sensitive demand. He shows that price regulation of NYC leads to inefficient matching because drivers making dynamic search decisions prefer searching locations with high profitability. Fixed pricing structure of taxis prevents the market from clearing on prices.

This paper follows the approach of Buchholz (2016) and extend his model. My contributions are twofold. First, I build a richer demand model that is not only sensitive to prices, but also sensitive to supply/demand to incorporate both INE and DNE. For instance, each geographic location can be deemed as a platform. Drivers' search decisions among locations are analogous to software providers' choosing platforms. High demand of passenger in one location attracts more drivers and more drivers

increase the choice probability of passengers. Frechette, Lizzeri and Salz (2016) also includes supply in passenger's demand function in the form of a simulated waiting time. But they do not model drivers' location choices. Second, I model cross-firm competition between taxi and Uber drivers for passengers. The first extension allows me to study non-price factors that influence matching efficiency. The second extension provides richer competition form between drivers and allows me to study regulation of Uber's supply on matching efficiencies of taxis, profits and consumer welfare. A very similar work to this paper is by Shapiro (2018) which however focuses on Uber's welfare contribution to the New York City without emphasizing the network effects in the matching efficiency.

This paper also contributes to empirical literature with dynamic oligopoly models. When there are a large number of firms within the market, Weintraub et al.(2007,2008) propose the concept of oblivious equilibrium(OE) to approximate Markov-perfect equilibrium in order to avoid the curse of dimensionality. In oblivious equilibrium, the firm is assumed to make the decision based only on its own state and deterministic average industry state rather than states of other competitors. In this paper, I assume drivers compete with the distribution of other drivers throughout the day. Under the OE assumption, only the distribution path at equilibrium is calculated. There are empirical papers using stationary OE (Xu(2008), Saeedi(2014)) and nonstationary OE (Qi(2013), Buchholz(2016)) to solve equilibrium of a model with large number of agents.

Finally, there are many works related to the taxi and ride-sharing industry. Some use the same trip records dataset as this paper. Early work studying NYC taxi industry include Farber (2005, 2008), Crawford and Meng (2011) which study taxy drivers' labor supply decisions. Frechette, Lizzeri and Salz (2016) study taxi drivers' labor supply decisions with matching frictions. In recent years, research on ride-sharing industry also grows. For example, Chen et al.(2017) study flexible labor supply of Uber drivers and Chen and Sheldon (2015) study surge pricing of Uber. Other work related to traffic condition and government regulation worth to mention is by Kreindler (2018) which studies road congestion pricing policy in India.

3 A One-Period, Two-Islands Model of Search and Matching

In this section, I build two simplified models to show the influence of network effects and duopoly competition on matching efficiency. These are the two main contributions of this paper to the literature. Both models are built under an environment of drivers searching for passengers among two islands in one period. The prices are fixed in both models. There are only taxis in the first model and there are taxis & Uber in the second. In the first model, I study how network effects influence matching efficiency by changing supply coefficient in the demand equation. In the second model, I study how competition and regulation affect matching by changing the total number of Uber cars. These exercises help to understand the mechanisms underling the dynamic structural model of this paper.

3.1 Monopoly Model

There are a fixed number of taxis, N_y, searching for passengers among two isolated islands $i = 1, 2$ in one period[7]. The fare in each island is denoted as p_i and is fixed. Supply and demand of each island is denoted as v_i and u_i. For simplification, I use a linear demand function in each island as a reduced form of aggregate demand over passengers' decisions[8]

$$u_i = -\alpha u_i + \beta v_i + d_i, \ \forall i \tag{3.1}$$

The coefficient $-\alpha$ on demand u_i, assumed negative, measures DNE of other passengers in the same island[9]. The coefficient β on supply, assumed positive, measures INE of supply on demand. The d_i captures island fixed effects such as population size and price. Since prices are fixed and the purpose of this exercise is not to model prices, I add prices to d_i. The goal of this exercise is to study how β affects equilibrium matching outcomes. A taxi chooses which island to serve in order to maximize his expected profit. The driver's optimization problem is:

[7] I use y as index for yellow taxis and x as index for UberX in this paper.

[8] One can think this demand equation as linear approximation for discrete choice model with one product and one outside option.

[9] The direct network effect could be positive or zero as well.

$$i^* = \arg\max_i \frac{m_i}{v_i} p_i \tag{3.2}$$

where m_i is matches in island i obtained from $m_i = \min\{u_i, v_i\}$ which means perfect matching within island. Assume all passengers and drivers make simultaneous decision and Nash equilibrium satisfies the conditions E1-E4:

$$\frac{m_1^*}{v_1^*} p_1 = \frac{m_2^*}{v_2^*} p_2 \tag{E1}$$

$$u_i^* = \frac{\beta}{1+\alpha} v_i^* + \frac{1}{1+\alpha} d_i \tag{E2}$$

$$m_i^* = \min\{u_i^*, v_i^*\} \tag{E3}$$

$$v_1^* + v_2^* = N_y \tag{E4}$$

Condition E1 means the expected profits of the two islands are equal and drivers have no incentive to deviate. Condition E2 means that demand u_i is calculated according to equation (3.1). E3 follows perfect matching assumption. Finally, E4 means the total number of taxis is fixed at N_y.

Proposition 1: There exists a Nash equilibrium such that one island exhibits excess demand and the other island exhibits excess supply. (see appendix for proof).

I consider one special equilibrium such that there is excess supply in one island(w.l.o.g. island 1) and excess demand in the other(island 2). The parameter values required for existence of this equilibrium are given in proposition 1. The equilibrium demand and supply are:

$$v_1^* = \frac{p_1}{p_2} u_1^* \qquad \text{(supply in island 1)}$$

$$u_1^* = \frac{d_1}{1+\alpha-\beta\frac{p_1}{p_2}} \qquad \text{(demand in island 1)}$$

$$v_2^* = N_y - v_1^* \qquad \text{(supply in island 2)}$$

$$u_2^* = \frac{\beta}{1+\alpha} v_2^* + \frac{d_2}{1+\alpha} \qquad \text{(demand in island 1)}$$

In this equilibrium, the price in island 1 is greater than price in island 2, $p_1 > p_2$.

The excess supply in island 1 is $v_1^* - u_1^*$ and excess demand in island 2 is $u_2^* - v_2^*$. The aggregate matching friction is measured as:

$$mismatch = \min\{v_1^* - u_1^*, u_2^* - v_2^*\} \tag{3.3}$$

Expression 3.3 is derived by subtracting taxis' aggregate pickups , $u_1^* + v_2^*$, from the minimum of total demand and supply, $\min\{u_1^* + u_2^*, v_1^* + v_2^*\}$. It counts the number of extra matches that could be made in an efficient matching. Our interest in this exercise is to do static comparative analysis of β on the friction $\min\{v_1^* - u_1^*, u_2^* - v_2^*\}$. However, one thing to notice is that (3.3) is a good measure of efficiency when both aggregate demand and supply are fixed. In the model, aggregate demand is not fixed. Thus, the discussion below is within certain domain.

The excess supply in island 1 is $v_1^* - u_1^* = (\frac{p_1}{p_2} - 1)\frac{d_1}{1 + \alpha - \beta\frac{p_1}{p_2}}$ which increases in β. The intuition is as follows: when β increases, demand is more sensitive to supply and therefore increases. Increased demand makes island 1 more profitable to drivers. However, in island 2 the increase in demand does not increase profit because matching probability is already 1 due to excess demand. Thus, some drivers switch to search island 1. In the new equilibrium, both demand and supply in island 1 increase proportionally such that $v_1^* = p_1/p_2 u_1^*$. The excess supply in island 1, $v_1^* - u_1^*$, increases.

The excess demand in island 2 also changes in the new equilibrium. Increasing β has two opposite effects on demand in island 2. First, it positively affects demand as coefficient on supply. Second, supply in island 2 decreases which in return decreases demand in island 2. The first order derivative of u_2^* w.r.t. β is $\frac{v_2^*}{1+\alpha} + \frac{\beta}{1+\alpha}\frac{dv_2^*}{d\beta}$. The first order derivative of excess demand $u_2^* - v_2^*$ w.r.t. β is $\frac{v_2^*}{1+\alpha} + (\frac{\beta}{1+\alpha} - 1)\frac{dv_2^*}{d\beta} > 0$ because $\frac{dv_2^*}{d\beta} < 0$ and $\beta < 1 + \alpha$[10]. Since both excess supply in island 1 and excess demand in island 2 increase in β, the matching friction becomes larger when β is larger. Thus, if a network effect was present ($\beta > 0$) in the real world, but ignored ($\beta = 0$) in the model, our measure of friction in equilibrium and policy simulation are incorrect. In the next subsection, I extend this model by adding Uber and show how supply competition affects matching friction at equilibrium.

[10] $\beta < 1 + \alpha$ is equilibrium condition as one can see in u_1^*. The denominator $1 + \alpha - \beta\frac{p_1}{p_2} > 0$. Due to $p_1 > p_2$, we must have $1 + \alpha > \beta$

3.2 Duopoly Model

In this subsection, I extend the monopoly model by adding Uber with fixed number of cars N_x. The demand function is modified such that there are cross elasticities. The new demand equation is:

$$u_{yi} = -\alpha u_{yi} + \beta v_{yi} + \underbrace{\gamma u_{xi} - \theta v_{xi} + d_{yi}}_{D_{yi}}, \forall i \tag{3.4.1}$$

$$u_{xi} = -\alpha u_{xi} + \beta v_{xi} + \underbrace{\gamma u_{yi} - \theta v_{yi} + d_{xi}}_{D_{xi}} \forall i \tag{3.4.2}$$

where y indicates taxi, x indicates Uber and i indicates island[11]. The parameters α and β measure DNE and INE of each firm as in previous model. The new parameters γ and θ measure cross elasticity of demand to the supply/demand of the other product. This equation can be deemed as linear approximation for discrete choice demand. I denote the last three terms as D_{yi} and D_{xi} for convenience in analogy to d_i in the monopoly model. However, D_{yi} and D_{xi} are endogenous here. The equilibrium conditions in this case are similar to monopoly model:

$$\frac{m^*_{f1}}{v^*_{f1}} p_{f1} = \frac{m^*_{f2}}{v^*_{f2}} p_{f2} \qquad \forall f = y, x \tag{E1}$$

$$u^*_{fi} = \frac{\beta}{1+\alpha} v^*_{fi} + \frac{D^*_{fi}}{1+\alpha} \qquad \forall i = 1, 2,\ f = y, x \tag{E2}$$

$$m^*_{fi} = \min\{u^*_{fi}, v^*_{fi}\} \qquad \forall i = 1, 2,\ f = y, x \tag{E3}$$

$$v^*_{f1} + v^*_{f2} = N_f \qquad \forall f = y, x \tag{E4}$$

The competition comes from D_{fi} such that opponent's supply decreases firm's demand and opponent's demand increases firm's demand. As mentioned above, I consider this relationship as linear form of a simple logit demand model with two products such that demand of one product depends on utilities of all products. In this exercise, I focus on an equilibrium with excess supply for taxis in island 1, excess demand for taxis in island 2 and excess supply of Uber in both islands. The goal of this exercise is to

[11] For the rest of this paper, I denote y as yellow taxis and x as UberX.

study how the supply of Uber N_2 affect taxis' matching friction. I choose equilibrium that Uber has excess supply in both island for two reasons. First, if Uber has excess demand in both islands, there will be multiple equilibria. Second, since I only study taxis friction in this exercise, it is not necessary to analyze equilibrium in which Uber also has friction (i.e. one island with excess demand and one with excess supply). Existence of such equilibrium is proved in proposition 2.

Proposition 2: There exists a Nash equilibrium such that one island exhibits excess demand and the other island exhibits excess supply for taxis. Both islands have excess supply of Uber (see appendix for proof).

To simplify the solutions, I impose one important assumption that $\gamma = 0$[12]. The competition still remains in θv_{fi}. Then, the equilibrium demand and supply satisfy:

$$v_{y1}^* = \frac{p_{y2}}{p_{y1}} u_{y1}^* = \frac{p_{y2}}{p_{y1}} \frac{-\theta v_{x1}^* + d_{y1}}{1 + \alpha - \beta p_{y1}/p_{y2}} \tag{3.5}$$

$$u_{y2}^* = \frac{\beta}{1+\alpha} v_{y2}^* + \frac{-\theta v_{x2}^* + d_{y2}}{1+\alpha} \tag{3.6}$$

$$v_{x1}^* = \frac{-\theta v_{y1}^* + d_{x1}}{-\theta N_y + d_{x1} + d_{x2}} N_x \tag{3.7}$$

$$v_{x2}^* = \frac{-\theta v_{y2}^* + d_{x2}}{-\theta N_y + d_{x1} + d_{x2}} N_x \tag{3.8}$$

Solving explicit form of demand and supply as function of parameters requires complex algebra and the intuition will be mixed. These equations look like best response functions in Cournot model and I will discuss the intuition based on these equations. When N_x decreases due to regulation, supply of Uber in both islands decrease according to (3.7)(3.8). From (3.5), decreasing v_{x1} will increase demand of taxis in island 1 which further increases taxis drivers' incentive to search island 1 and therefore v_{y1} increases. More taxis in island 1 further decreases supply of Uber in island 1 as shown in (3.7). Due to the proportional increases of demand and supply of taxis in island 1, excess supply of taxis in island 1 increases. Since more taxis switch to island 1, taxi supply v_{y2} in island 2 decreases. Supply of Uber in island 2 has ambiguous change. The downside force to Uber's supply in island 2 is drop of N_x. The upside force is decreased supply of taxis in island 2 and switch of Uber drivers from island 1 to island 2. Supposing that

[12] I also impose $p_{x1} = p_{x2}$ which is irrelevant to study taxis' friction .

v_{x2} decreases, derivative of $u^*_{y2} - v^*_{y2}$ w.r.t. N_x ,that is $(\frac{\beta}{1+\alpha} - 1)\frac{dv^*_{y2}}{dN_x} - \frac{\theta}{1+\alpha}\frac{dv^*_{x2}}{dN_x}$, is negative. In this case, both excess demand and excess supply of taxis increase after N_x decreases. It implies matching efficiency is worse off under supply regulation of Uber. However if v_{x2} increases, it offsets the effect of decreased taxi supply in island 2. Whether excess demand of taxis increases or decreases is ambiguous. So does aggregate matching friction of taxis in this case.

To summarize, in this section, I build two simple spatial search models illustrating two findings. First, the interdependence of demand and supply in search and matching process (network effects) affects aggregate matching friction. Second, in a duopoly model, competition also affects the matching friction of each firm. Especially, when decreasing the supply of Uber, taxis' aggregate matching friction may increase or decrease depending on elasticities of demand. In the next section, I introduce the background of NYC taxi industry in which I empirically study matching efficiencies in a fully developed dynamic version of the search model.

4 Industry Background

My model application is based on the New York City taxi industry. In NYC, there are mainly two ways to get a ride, taxi or for-hire-vehicle(FHV). Taxis can only pick up street hails and FHV can only pick up pre-arranged ride requests. These two markets are strictly separated under the regulation of NYC government. Running a taxi requires a medallion attached to the vehicle. The total number of medallions available is fixed by regulation which is 13,587 in 2015. Medallion owners can trade medallions through auction. Along with yellow taxis, there are 7,676 boro taxis introduced to the city in 2013. Boro taxis follow the same rules as yellow taxis except for that they can only pick up passengers in Northern Manhattan, the Bronx, Brooklyn, Queens and Staten Island. Moreover, the boro taxis can only pick up passengers at the airport that prearranged. Yellow and boro taxis follow the same pricing rule under the regulation of Taxi and Limousine Commission (TLC). The medallion owner can either operate the taxi himself or lease to other licensed drivers. In 2015, there are 38,319 active taxi drivers running 13,587 vehicles. Usually, one driver operates the vehicle for a shift of the day. The day shift starts at 6 a.m. and night shift starts at 4 p.m. such that the expected revenues are equal between shifts. Part of the medallions are owned by

individual owners and part are owned by fleets. Regardless of medallion ownership, the operation of the vehicle is by an individual driver who either owns or leases the car.

The FHV also has different types, black car, livery or luxury limousine. Only black cars can provide contracted service through a smartphone app. Other types of FHV can only provide for-hire service by pre-arrangement. All FHV vehicles are required to be affiliated with black car, livery or limousine bases. Unlike taxis which have a fixed number of medallions FHV is allowed for entry. Black cars include ridesharing companies such as Uber, Lyft and Via[13]. Uber is a technology firm that provides a mobile app which creates a two-sided market for on-demand transportation. Riders send a request for a ride to Uber drivers through the Uber app. The information provided in the mobile app includes fare (calculated on distance and time of the trip), and waiting time before passengers are picked up. Active Uber drivers nearby receive the request and they can choose either to take the order or not[14]. However, Uber drivers do not know the destination before accept the request. If one driver doesn't take the request, it will be forwarded to another driver and so on. When demand for Uber is high but supply is low, Uber charges passengers the regular fare multiplied by a surge multiplier. By raising the fare, it intends to attract more Uber drivers to compensate the demand and supply gap.

Unlike taxis, most Uber drivers work part time and use their own cars to provide ride services. This makes it difficult to study Uber supply without proprietary data that I will discuss later. Uber also provides different services such as UberX, UberTaxi, UberPool, UberXL, SUV, etc. In this paper, I do not distinguish car types of Uber. I treat all trips completed by Uber cars affiliated with black car bases as identical.

5 Data

The data used in this paper comes from three sources. The main information about taxis and Uber cars comes from trip records provided by the New York City Taxi and

[13]Uber and Lyft also have black car bases. UberX drivers need to be affiliated with one of Uber black car bases in NYC. However, Uber vehicles are not physically dispatched from the bases such as luxury cars.

[14]An active Uber driver means that a driver opens Uber app and is willing to pick up passengers.

Limousine Commission (TLC)[15]. The taxis trip records include all trips completed by yellow taxis since 2009 and by boro taxis since 2013. Each trip is an observation in the data including pick-up and drop-off date/time, geographic location, trip distance and fare. One can calculate the interval between pick-up and drop-off time to figure out trip time. One important limitation of the data is that there is no identifier of the taxi vehicle for each trip. Moreover, I cannot tell where the vacant taxis are located until they pick passengers up. Because of this, supply of taxis at any time-location is not directly observed. Similarly, given only pickups data, I can not tell how many passengers who want rides fail to match a taxi .

The other part of TLC data is FHV trip records which include trips by black cars and luxury limousine. The method of collecting FHV data is different from taxis which is submitted by FHV bases. Each observation in the FHV data includes the base id that dispatches the vehicle for this trip, pick-up date/time and taxi zone area of the pick-up[16]. I obtain the Uber trips according to base numbers that the vehicle is affiliated with. The qualities of the data submitted by bases also differ across companies. For example, only trip records submitted by Uber bases include the pick-up zone. Thus, I only model Uber as competitor to taxis without Lyft as a player. Lyft is not an effective player at that time. Out of all black car trips, Uber accounts for 72.6% and Lyft accounts for 11.6%. Unlike the taxi records, drop-off, trip distance, fare and trip time are not observed for Uber. I solve the problem of trip time and distance by assuming the travel distance and time between two locations are the same as taxis' trips between the same locations at the same time. The fare of Uber is calculated according to Uber's price rule after knowing travel time and distance. The drop off locations of Uber are generated by my structural model.

The second source of data that supplements the main trip record is Uber's surge multiplier. Uber's fare during a normal time is calculated based on trip distance and time. However, during rush hours or when Uber's supply is less than demand, Uber applies surge pricing which multiplies the regular fare by a surge multiplier. To calculate Uber's trip fare, I use Uber's API for developer to collect the real time surge multiplier every 10 minutes at 79 selected location spots across the city from November 2015 to June 2016. Each request via the API returns the surge multiplier

[15]In the past years, the data is available to public by filing FOIL request to TLC. Now, all trip records are accessible from TLC's website, http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

[16]The taxi zones are not accurate as geographic locations which are areas defined by the TLC. There are about 263 taxi zones in the NYC.

of different types of Uber products at that time and I use multiplier of UberX as representative. Combining with trip records data by matching location and time, I can calculate approximate fare of Uber trips.

The third data I use is subway riderships obtained from Metropolitan Transportation Authority (MTA) of the city. This data is used to calculate the number of potential riders of a given location-time, a measure of market size. The ridership data includes information on weekly aggregate entrances to each station of the NYC subway. For a given station, the riderships are sorted by various types of MetroCards that the customers swipe such as 30 day pass, student, and full fare. I only count those paying full fare as potential passengers of taxis and Uber since they are more likely to have the same travelling patterns as taxi&Uber passengers compared to commuters. Thus, the market size of a location at a given time is defined as the sum of taxi and Uber pickups and full fare riderships. Those who choose subway as the outside option and who fail to match a car comprise full fare riderships. To divide weekly aggregate subway riderships, first I allocate ridership evenly to all locations near the station, then I divide subway riderships of each location evenly for 7 days, and finally I proportionally divide the daily riderships of a location based on taxi and Uber pickups distribution over time of day.

5.1 Sample Construction

In the empirical part of this paper, I model drivers' dynamic search across locations over daytime of a representative weekday. I choose April of 2016 as my sample period. My model focuses on equilibrium evolution of pickups over a day and therefore for a given location-time I average pickups over all weekdays of April 2016 as steady state pickups of this market[17]. The time period within a day in my sample is restricted to 6 a.m. to 4 p.m. Hence, my empirical model studies search and matching during 6 a.m.-4 p.m. of a representative weekday of April 2016. The main reason for doing this is that I do not have information about how many active Uber drivers at a given time of day. Uber drivers have much more flexible work time than taxis due to free entry

[17] I do not model daily search and matching since computing (unobserved) supply and demand of all days requires solving dynamic equilibrium many times which is computationally costly. In the data, I find that pickups distributions over all weekdays are quite similar and by averaging across all weekdays I have good approximation of how pickups evolve within a representative day.

and exit[18] . Given that I have no real-time data on active Uber drivers, I cannot model weekends and night shift of weekdays when part-time Uber drivers are more likely to be active. The implicit assumption I make is that during 6 a.m.-4 p.m. of weekdays, the number of active Uber drivers is fixed (full-time drivers) such that supply is tractable. The time period also covers the day shift of taxis such that the number of taxis is also fixed.

I discretize time and space in the following ways. I define every 10 minutes as a time period and 60 periods in total. Within each period, passengers and drivers randomly meet only once within a location and successful contacts become pickups of that location-period. In other words, each driver only supplies once in a period[19]. I divide the city and select 40 geographic locations as shown in figure 1. I define the area of each market by combining small taxi zones and comparing pickups. For example, the area sizes of locations in Queens and Brooklyn are large compared to those in Manhattan because the pickups in outer boroughs are quite less. I exclude central park from this map since all pickups within it is on the boundaries of the park and I assign pickups in central park to locations nearby. The combination of location-period pair is defined as a market (platform) in this paper. Finally, the pickups of a market are calculated as monthly average pickups of the same market over all weekdays in Aprial 2016.

The variables constructed from the data include trip distance, trip time, fares and trip distribution. In a period, for a trip between any two locations, I calculate monthly average trip distance and trip time over all taxis' trips between the same two locations and of the same period[20]. I use this average trip distance and time of any given origin-destination-period to calculate average fares of taxis and Uber using their respective pricing structures. I multiply Uber's regular fare by monthly average surge multiplier of the same origin-period to calculate the final Uber fare. The location spots I choose to collect the surge multiplier are shown in figure 1. The trip time in minute is transformed into number of 10-minute periods. For example, a 25 minutes trip takes 3 periods to complete.

[18]One can check studies on labor supply by Chen et al.(2017) and Hall and Krueger (2016) for Uber and by Farber(2008) and Frechette et al. (2016) for taxi

[19]Under this assumption that each driver only supplies once in a period the model could underestimate supply if the length of the defined period is long. For example, drivers could complete a trip within 10 minutes and pick up another passenger.

[20]This can only be calculated from taxis' trips since data of Uber dropoffs is not available. I assume the trip distance and time are same for Uber.

For any well defined market as a location-period combination, I construct market size and distribution of passengers' destinations. Market size is widely used in discrete choice demand model to control substitution among inside products to outside option when price increases. In my demand model, the outside option of demand is subway. The population of a market is calculated as sum of subway riderships and taxis & Uber pickups. Note that, these subway riderships include both travellers who choose subway when making discrete choice and those who choose taxis & Uber but fail to match one. I calculate the market size in the following way. First, I divide weekly aggregate riderships of a station by seven as daily average riderships. Many stations are located at intersection corner of locations. I evenly assign riderships of a station to nearby locations. Then I assume the subway riderships follow the same trendency of taxis & Uber pickups over the time of day and divide subway riderships proportionally over time of day[21].

Finally, I calculate the trip pattern of all travellers. I do not have data of distribution of all passengers' destinations. Instead, I calculate the dropoff pattern of taxis in November 2010 as proxy for population travelling pattern[22]. There are two implicit assumptions in order to use this pattern as population transportation pattern. First, I assume travellers paying full fare for subway follow the same travelling pattern to taxis' passengers in 2010. Second, the travelling pattern of 2010 and 2016 are the same after Uber's entry. The trip pattern is defined for each market as shares of destinations. In other words, taxis' distribution of dropoffs in 2010 is deemed as travelling pattern of population in 2016 and distribution of taxis' dropoffs in 2016 is outcomes of travellers' discrete choice demand. In the next subsection, the sample and data overview are provided.

5.2 Sample Overview

The table 1 shows monthly aggregate statistics of weekday pickups in November 2010 (22 days) and April 2016 (21 days). It demonstrates the distribution of pickups by area, firm and shift. In April 2016, taxis' monthly aggregate pickups during day shifts is 3.7 million. Almost 93% of the total pickups are in Manhattan, 4.59% are in JFK and Laguardia airports. My sample of 40 locations cover 99.38% of all taxis' pickups

[21]By doing this, I retain the variations of subway market share across locations but not over time of day.

[22]The reason of choosing November 2010 is that Uber and boro taxis are not available.

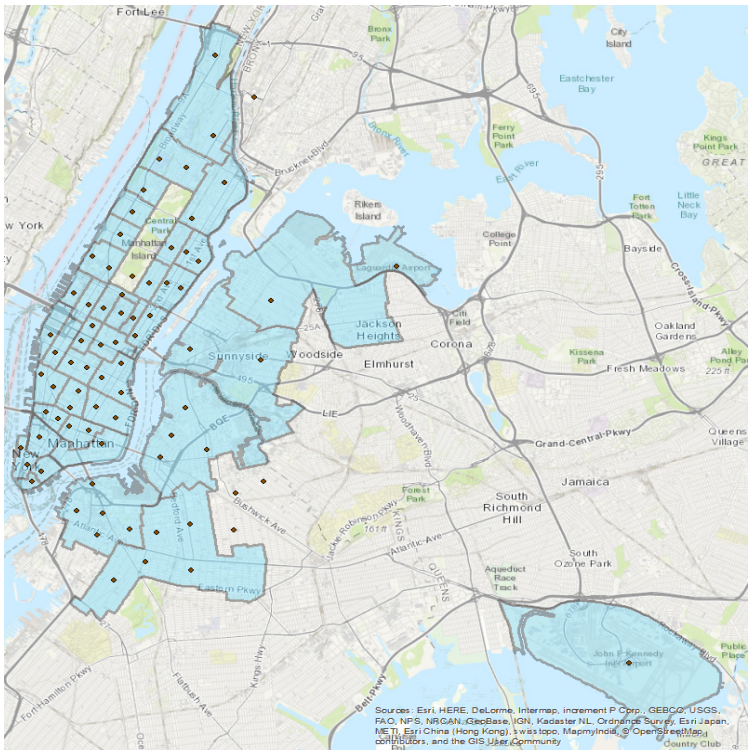


Figure 1: 40 select markets and 79 spots to collect Uber surge multiplier

during day shifts. Outer boroughs has 2.24% pickups in total. Comparing pickups of taxis between 2010 and 2016, I can observe huge decrease of pickups from 4.6 million to 3.7 million for day shifts. The pickups distribution also has small differences that share of airport increases from 3.42% to 4.59% and share of Manhattan drops from 93.67% to 93.17%. More differences can be investigated if I collapse Manhattan in many smaller zones and compare the shares of pickups. This indicates that taxis' supply and demand pattern changes slightly after Uber's entry. Table 1 also includes Uber's and Lyft's pickups distribution in April 2016. Uber has different distribution in comparison to taxis that 59.5% pickups are in Manhattan. The share of Uber's pickups in outer boroughs is 36.58% quite larger than taxis. It indicates that during my sample period, only half of Uber's pickups have direct competition with taxis. The share of Uber's pickups covered by my 40 locations is 77.63%. Lyft as second largest FHV firm has 0.2 million pickups far less than Uber. Similarly, most of Lyft pickups are in outer boroughs. In my demand model, I exclude Lyft from choice set. The variation of pickup shares of taxi and Uber across locations as shown in table 1 helps estimate demand model similar to BLP demand model, in which variation of market

shares among products both within the same market and cross markets helps identify price elasticities, product fixed effects and mean utility. The complexity in my model is that pickup shares are not exactly demand shares considering mismatches within market.

In random coefficient discrete choice model, we use market demographics to identify random price coefficient. In my model, the demographics come from exogenous travelling patterns of passengers defined by distribution of destinations conditional on any given market. I use taxis' dropoffs data of November 2010 to approximate market demographics. I do not use dropoffs in 2016 as market demographics because it is endogenous outcomes of passengers' discrete choice among taxi, Uber and outside option. Table 2 provides a rough overview of distribution of dropoffs. The distribution is calculated using pickups and dropoffs in day shift of weekdays. The first panel tells that 95.11% of pickups in Manhattan are delivered within Manhattan and 3% to airports. Trips originating from airports have 73.45% ending up in Manhattan and 6.37% of them are inter-airport. Comparing 2010 and 2016, the dropoff distributions are slightly different that trips originating from airports to Manhattan decrease from 73.45% to 71%. Table 2 only shows travelling patterns among three highly aggregate areas, Manhattan, Airport and Other. More variations of travelling patterns can be discovered at the market level, which helps identify random coefficient of prices. For instance, given two markets with same prices, market size and supplies, differences in demands reflect differences in travelling patterns. Finally, the destination distribution of taxis in 2016 also helps with my demand estimation such that model predicted dropoffs of taxis match observations in the data.

Table 1: Trip and Share by Firm, Shift, and Area

Firm&Shift	*Total*	*Manhattan*	*Airports*	*Other*	*40 mkt*
Yellow Taxi 2010.11(22)					
Day shift	4,627,258	93.67%	3.42%	2.91%	98.93%
Night shift	5,139,146	92.91%	3.46%	3.63%	98.99%
Yellow Taxi 2016.04(21)					
Day shift	3,730,326	93.17%	4.59%	2.24%	99.38%
Night shift	4,279,262	92.04%	4.94%	3.02%	99.27%
Uber 2016.04					
Day shift	1,322,507	59.5%	3.92%	36.58%	77.63%
Night shift	1,989,054	64.15%	4.1%	31.75%	82.43%
Lyft 2016.04					
Day shift	215,240	40.10%	2.46%	57.44%	59.56%
Night shift	301,577	49.64%	2.84%	47.52%	40.44%

Table 2: Distribution of Dropoffs in Day Shift by Firm

	Obs.	*Manhattan*	*Airports*	*Queen&Brooklyn*	*not in 40*
Yellow Taxi 2010.11(22)					
Manhattan	4,334,266	95.11%	3%	0.98%	0.89%
Airports	158,410	73.45%	6.37%	8.15%	12.01%
Queen&Brooklyn	84,925	47.53%	4.43%	42.13%	5.88%
not in 40	49,657	41.50%	3.2%	9.04%	46.24%
Yellow Taxi 2016.04(21)					
Manhattan	3,475,467	94%	3%	0.97%	1.17%
Airports	171,407	71%	3.2%	10.73%	14.69%
Queen&Brooklyn	60,274	42.6%	4.26%	44.89%	8.2%
not in 40	23,178	45%	3.9%	17.3%	32.8%
Uber 2016.04					
Manhattan	786,854				
Airports	51,855				
Queen&Brooklyn	187,912				
not in 40	295,886				

At last, table 3 shows the statistics of key variables in my sample construction discussed in section 5.1. Given 40 locations and 60 periods in my sample, there are 2,400 well defined markets. Uber's surge multiplier varies from 1 to 1.37 as 90% quantile. Taxi and Uber prices are calculated at origin-destination-time level which has 96,000 observations (i.e. 40 destinations of each market). Uber's fare on average is higher than taxi for the following reasons: 1, Uber charges both trip time and distance; 2, there is surge multiplier; 3, Uber charges minimum fare $7 which is higher than taxis for short trips.

Table 3: Summary Statistics of Key Variables

variable	*Obs*	*mean*	*10%ile*	*90%ile*	*S.D.*
trip distance	3,643,011	2.61	0.6	5.6	3.34
trip time	3,636,906	15.1	4.4	29.38	11.63
trip fare	3,643,011	12.53	5	23.5	9.6
final sample variables					
surge	2,400	1.14	1	1.37	0.18
taxi matches	2,400	72.28	8.80	163.02	57.34
Uber matches	2,400	20.37	8.33	34.95	10.25
taxi fare	96,000	17.68	7.92	28.4	10
Uber fare	96,000	22.39	9.12	37.95	12.48
trip distance	96,000	5.07	1.19	9.88	4.10
trip time	96,000	23.78	8.97	39.11	12.35

In figure 2, it displays the comparison of trip time in 2010 and 2016. The difference in distributions implies worse traffic condition in 2016. For example, the median of trip time in 2010 is 18.73 minutes. The median of trip time in 2016 is 22.62 minutes. The trip time in terms of periods of 2016 is also higher than 2010. In 2010, more than half of trips take less than two periods to deliver. However, more than half of trips in 2016 takes 3 or more periods to deliver. In the second counterfactual of this paper, I replace the trip time of my sample by the trip time of 2010 to study how traffic improvement affects the matching efficiency in new equilibrium.

In order to show the differences of expected profits across markets, I calculate the expected profit conditioning on randomly picking up a passenger for each market. I obtain 2400 ($40 * 60$) expected profits in total. The distributions of expected profits for taxi and Uber are shown in figure 3. For taxis, the average expected profit is 15.27 and it ranges from less than 10 dollors to almost 60 dollors. The distribution for Uber has

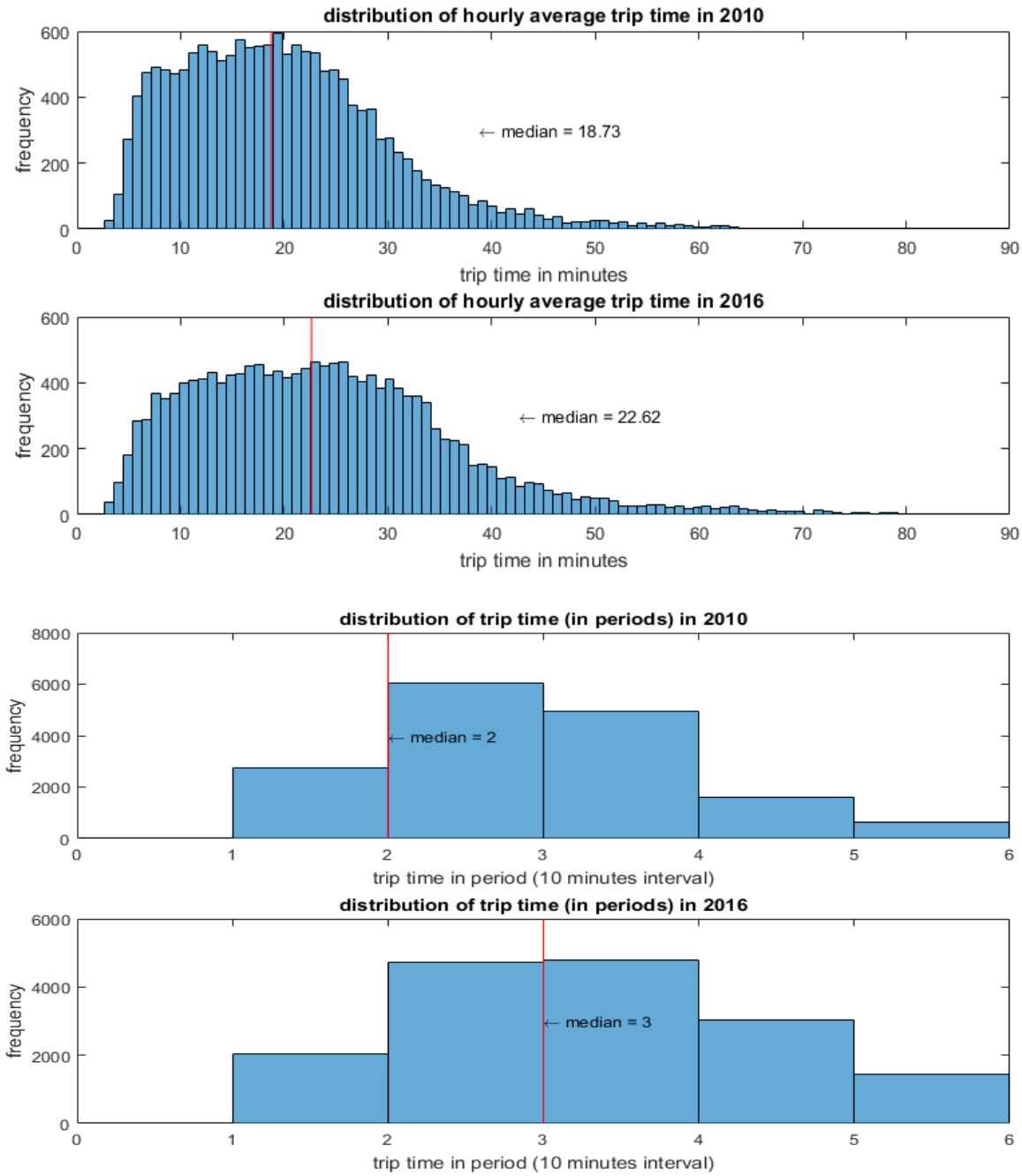


Figure 2: Comparison of traffic speed in 2010 and 2016

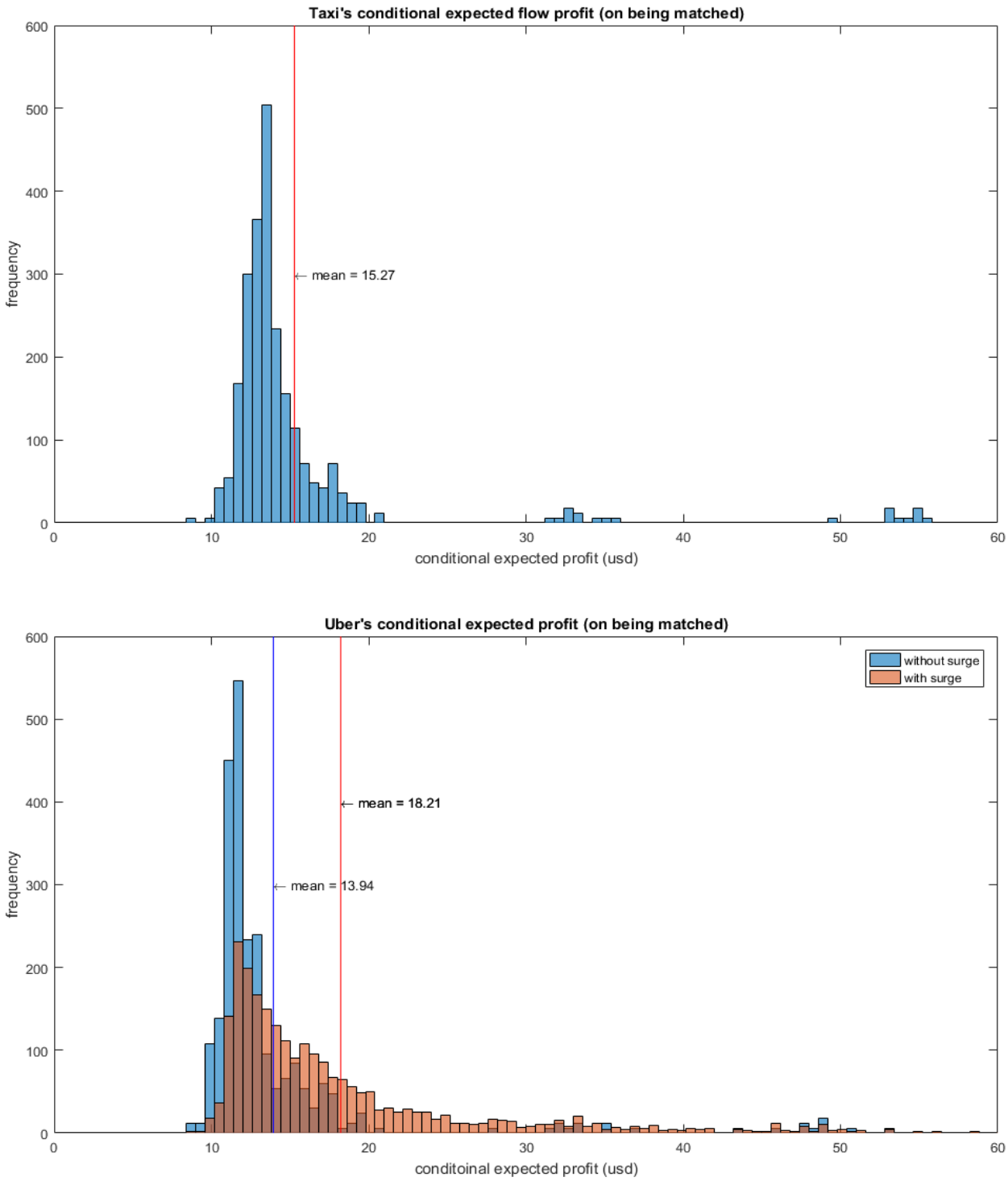


Figure 3: Distribution of expected profits of taxi and Uber

mean of 13.94 dollars without surge multiplier and 18.21 dollars with surge multiplier. It implies that Uber driver's expected profit is higher than taxi in general. This high heterogeneity in profitability as shown in figure 3 illustrates the incentive of drivers' search decisions and why some markets are oversupplied than others. However, the expected profit calculated here is static flow profit. Drivers' search decisions are made based on dynamic search values. In the following sections, I will build a dynamic model and with the estimates I can compare the differences in search values across markets.

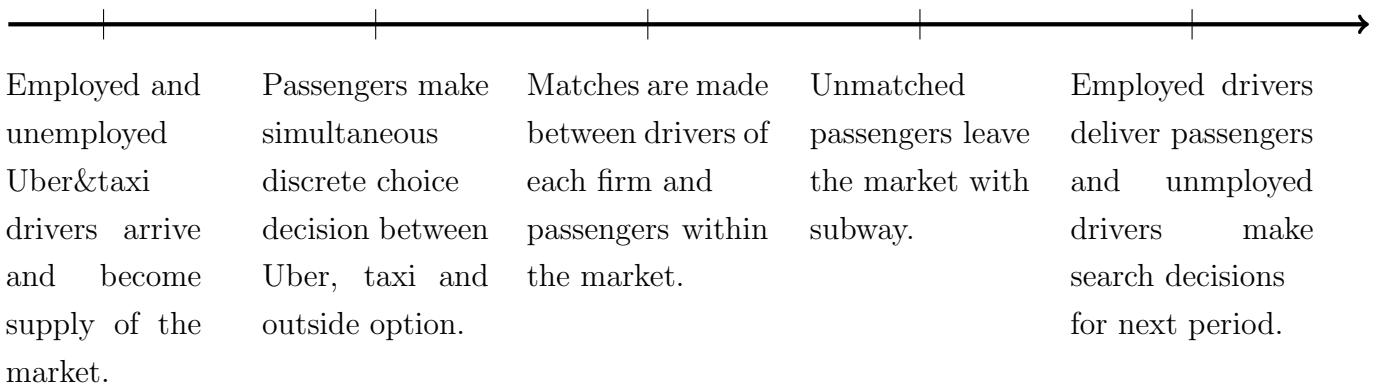
6 Empirical Model

The structural model fully extends the search and matching model discussed in section 3. Taxi and Uber drivers make dynamic spatial search decisions among I locations over T periods in a day. Potential passengers make static discrete choice decision among Uber, taxi and subway. Drivers and passengers have perfect information about the size of either side in a given market. When making supply/demand decision, the agent accounts for both the indirect network effects from the other side of the market and direct network effect from the same side. This supply sensitive demand specification is one contribution of this paper relative to Buchholz (2016). The model allows two types of frictions that prevent the market from clearing. First, within location I allow taxis and passengers to not fully contact with each other due to coordination failure as in Burdett, Shi and Wright (2001). However, I assume perfect matching of Uber within the market. Matching is perfect in airports for both taxis and Uber. In other words, I simplify the matching process within a market by assuming a matching process with an explicit functional form. Second, across locations because of drivers' endogenous search decisions there are locations exhibiting excess supply along with other locations with excess demand. Both frictions result in inefficient matching at the city aggregate level.

Before delving into demand and supply decisions, I assume the timeline within a market is as follows. At the beginning of each period, part of taxis and Uber cars will arrive at their destinations. If the car has a passenger on board (employed), it arrives at the dropoff location. If the car is vacant (unemployed), it arrives at the location based on the driver's search decision in the last decision period. Some of the cars either employed or unemployed are still on their way to the destinations and will not necessarily arrive at a location in this period. All arriving cars become supply to

that market in this period[23]. A passenger in this market has perfect information about fares and beliefs on demand/supply, how likely he will find a taxi or Uber car, and how long it takes to match[24]. Passengers make static discrete choice decision. Aggregating all passengers' decisions returns demand for each firm in this market. Then matches are made within market and firm. Unmatched passengers either due to excess demand or matching friction leave with subway. Employed drivers deliver passengers to their destinations and unemployed drivers choose next locations to search.

timing of supply, demand and match of a market



6.1 Passengers' Choice Problem

In a market defined by a location-period combination, a group of potential travellers (market size) make discrete choice among taxis, Uber and subway conditional on their exogenous destination with knowledge on prices, product qualities, supply and demand. In this model, Uber is denoted as x (UberX), taxi as y (yellow taxi) and outside option as o. The utility of a passenger c in location i at period t choosing firm $f = x, y, o$ to travel to j prior to matching process is:

[23]It assumes that a driver must stay in this location for at least one period.

[24]Uber's supply can be perfectly learned by app which shows how many cars around and how long to wait. Taxis' supply is hard to directly observe. However, my model studies demand and supply only in equilibrium such that passengers are fully experienced and know how likely to get a car without necessarily knowing how many cars nearby.

$$
\begin{aligned}
U^{ij}_{cft,pre} &= G(\tau_f(u^i_{ft}, v^i_{ft}), u^i_{ft}, v^i_{ft}, p^{ij}_{ft}, X^i_{ft}, \varepsilon^i_{cft}) \\
&= \tau^i_{ft} U^{ij}_{cft,post} + (1 - \tau^i_{ft}) U^{ij}_{cot,post}
\end{aligned} \tag{6.1}
$$

where u^i_{ft} is market demand for firm f, v^i_{ft} is firm f's supply and p^{ij}_{ft} is the price from i to j. The function $\tau^i_{ft} = \tau_f(u^i_{ft}, v^i_{ft})$ is the probability of being matched by choosing firm f determined by the firm specific demand and supply level in the market. The matching probability does not differ for different destinations j under the implicit assumption that drivers cannot reject a ride. The probability $\tau(u^i_{ft}, v^i_{ft})$ can be written as $m(v^i_{ft}, u^i_{ft})/u^i_{ft}$ with matching function $m(v^i_{ft}, u^i_{ft})$. The function form of m will be discussed later. In addition to effects of u^i_{ft}, v^i_{ft} on matching probability, they could also affect the utility through classical direct network effects. For example, the user base u^i_{ft} in the market could affect individual's choice decision as outcome of consumers learning from each other or herd behaviors (Goolsbee and Klenow (2002)). The supply v^i_{ft} affects choices through indirect network effect[25]. Higher v^i_{ft} could increase utility by decreasing waiting time after conditioning on being matched. I assume $U_{cot,post} = 0$ which implies that unmatched passengers end up with zero utility by taking subway. Furthermore, I assume the logarithm of $U^{ij}_{cft,post}$ is linear such that (6.1) can be rewritten as :

$$
\begin{aligned}
\ln(U^{ij}_{cft,pre}) &= \ln(\tau^i_{ft}) + \ln(U^{ij}_{cft,post}) \\
&= \underbrace{\theta_1 \ln(v^i_{ft}) + \theta_2 \ln(u^i_{ft}) + d_x + d_i + t + \xi^i_{ft}}_{\delta^i_{ft}} + \alpha^{ij} \ln(p^{ij}_{ft}) + \varepsilon^{ij}_{cft}
\end{aligned} \tag{6.2}
$$

Equation (6.2) is obtained by: 1, transforming τ^i_{ft} as a linear combination of $\ln u^i_{ft}, \ln v^i_{ft}$; 2, assuming $\ln(U^{ij}_{cft,post})$ is linear in $\ln u^i_{ft}, \ln v^i_{ft}$ and other charateristics including price p^{ij}_{ft}, Uber fixed effect d_x, market fixed effect d_i, time fixed effect d_t, unobserved demand shock ξ^i_{ft} and idiosyncratic shock ε^{ij}_{cft}. The benefit of these assumption is that all endogenous variables of the model u^i_{ft}, v^i_{ft} are contained in parameter δ^i_{ft} such that demand is simple to solve. In other words, unobserved endogenous demand u and supply v which need to be solved through structure are separated from estimating price

[25] In this paper, I misuse the concepts of indirect network effect and cross-network effect so that they are interchangeable.

coefficients[26]. The drawback of log-linearity assumption of τ is that coefficients θ in (6.2) measures the joint effect of v^i_{ft} or u^i_{ft} without distinguishing channels through matching probability or through classic network effects (i.e. product variety, word-of-mouth.). The price coefficients α^{ij} depends on travel distance and trip type which are parameterized as:

$$\alpha^{ij} = \sum_{k=1,2,3} \alpha_k \mathbb{1}\{dist^{ij} \in \mathcal{I}_k\} + \alpha_4 \mathbb{1}\{dist^{ij} \in \mathcal{I}_{JFK}\} \tag{6.3}$$

where $\mathcal{I}_1$ is for trip distance less than 3 miles, $\mathcal{I}_2$ is for distance between 3 and 6 miles, $\mathcal{I}_3$ is for distance greater than 6 miles and $\mathcal{I}_{JFK}$ is trips between JFK airport and Manhattan which charges flat rate. Finally, I assume the utility of choosing subway before the matching process as:

$$\ln(U^{ij}_{cot,pre}) = \delta^i_{ot} + \varepsilon^{ij}_{c0t} \tag{6.4}$$

where δ^i_{ot} is normalized to zero. Since the subway fare is fixed for single trip regardless of trip length, there is not price in 6.4.

I allow substitution between taxi and Uber by assuming a nested logit demand model such that:

$$\varepsilon^{ij}_{cft} = \zeta^{ij}_{cgt} + (1-\beta)\nu^{ij}_{cft} \tag{6.5}$$

where ζ^{ij}_{cgt} is common to taxi and Uber which are categorized as one group, and subway alone as the other group. Variable ν^{ij}_{cft} is assumed to follow type I extreme value distribution. The distribution of ζ^{ij}_{cgt} satisfies that ε^{ij}_{cft} is also an extreme value random variable. The parameter β measures substitution between taxi and Uber. When $\beta = 0$, it is equivalent to the simple logit demand model. Larger β implies stronger substitution pattern between taxi and Uber. Then, the choice probability conditioning on route ij at time t becomes product of choice probability within group and probability across group. Within the group of taxi and Uber, the choice probability is:

[26]In details, given values δ^i_{ft} and α^{ij}, demands are fixed when I iteratively solve equilibrium supply. If matching probabilities interacts with price, update of supply requires update of demand as well. More details are in section 7.

$$s^{ij}_{y|gt} = \frac{\exp((\delta^i_{ft} + \alpha^{ij}\ln(p^{ij}_{ft}))/(1-\beta))}{\exp((\delta^i_{yt} + \alpha^{ij}\ln(p^{ij}_{yt}))/(1-\beta)) + \exp((\delta^i_{xt} + \alpha^{ij}\ln(p^{ij}_{xt}))/(1-\beta))} \tag{6.6}$$

Let us denote the denominator of equation 6.6 as:

$$D_g = \exp((\delta^i_{yt} + \alpha^{ij}\ln(p^{ij}_{yt}))/(1-\beta)) + \exp((\delta^i_{xt} + \alpha^{ij}\ln(p^{ij}_{xt}))/(1-\beta)) \tag{6.7}$$

The probability of choosing the group of taxi and Uber is:

$$s^{ij}_{gt} = \frac{D_g^{1-\beta}}{1 + D_g^{1-\beta}} \tag{6.8}$$

Then the choice probability becomes $s^{ij}_{ft} = s^{ij}_{f|gt} * s^{ij}_{gt}$.

In the traditional way, I can estimate demand by matching choice probabilities of to the market shares obtained by dividing demand u^{ij}_{ft} by the size of people travelling from i to j as Berry (1994). However, there are two obstacles to do this in this paper. First, I can only observe pickups m^{ij}_{ft} rather than demand u^{ij}_{ft}. Thus, I cannot calculate market share of demand. Second, even though assuming matches equal to demand such that $u^i_{ft} = m^i_{ft}$, I cannot calculate u^{ij}_{xt} for any j of Uber without knowing the destination distribution of Uber trips. In other words, I can not directly estimate equation (6.2) at the route level $\{i, j, t\}$. Instead, I treat the choice probability as the model prediction for the conditional (on destination) market share and aggregate s^{ij}_{ft} over j to calculate the unconditional market share.

The exogenous distribution of passengers' destination in a market $\{i, t\}$ is denoted as $A^i_t = \{a^{ij}_t\}_{\forall j}$ where a^{ij}_t is the probability that a passenger from this market travels to j. The unconditional market share predicted by the demand model is:

$$s^i_{ft} = \sum_j a^{ij}_t s^{ij}_{ft} \tag{6.9}$$

Denote the market size as λ^i_t. I can calculate the potential demand before matching process as:

$$u^i_{ft} = \lambda^i_t s^i_{ft} \tag{6.10}$$

In comparison to the exogenous distribution of travellers' destinations A_t^i, I can also calculate the dropoffs distribution of each firm $\tilde{A}_{ft}^i$ as outcomes of passengers' discrete choice using Bayes' rule. Thus, the model predicted firm specific destination distribution becomes:

$$\tilde{a}_{ft}^{ij} = \frac{a_t^{ij} s_{ft}^{ij}}{s_{ft}^i} \tag{6.11}$$

I put tilde and firm index f in $\tilde{a}_{ft}^{ij}$ to distinguish from a_t^{ij}.

To summarize the demand side, I assume passengers make demand decision before matching process but with belief of demand and supply level as proxy for matching probability, waiting time, and network effects. Given a set of demand parameter values, the demand model can predict two main things. First, the model predicts market shares s_{ft}^i and demand u_{ft}^i. Second, it predicts endogenous distribution of firm's dropoffs $\tilde{\boldsymbol{A}}_{\boldsymbol{f}}$. Though, I cannot directly observe demand and supply in the data, in the estimation section, I will discuss how to solve supply/demand by fitting model predicted pickups to pickups observed in the data.

6.2 Drivers' Choice Problem

At the end of each period, if the driver is employed, he will travel to the destination requested by the passengers. Drivers can not refuse to deliver a passenger once being matched. The probability of an employed car of firm f in location i at time t travelling to destination j is $\tilde{a}_{ft}^{ij}$ which is obtained from equation (6.11). Search decisions are made only by unmatched drivers at the end of each period.

If the driver is unmatched after the current period, he makes a decision on which location to search for passengers in the next period. Drivers are identical within firm and make individual decisions without coordination of the firm. Similar to passengers, when drivers consider a location to search in the next period, they know the matching probability, expected profit conditional on being matched and continuation value if not matched in that location. In order to know the matching probability, drivers need to have rational expectation of demand and supply distribution across locations in the future. A standard dynamic oligopoly model is inappropriate for this game due to the large number of drivers in the game. For example, the number of possible states of

allocating N drivers into I locations will be C^{I-1}_{N+I-1} which is large when N is large[27]. The model would be intractable and computationally infeasible if drivers' expected profits are taken over all possible market states. Instead, I assume drivers make search decision only on their own state and knowledge of the deterministic market evolution of demand and supply distributions. This concept comes from oblivious equilibrium (Weintraub et al.(2008)) when players are atomistic and individual decision does not measurably impact aggregate market state. In equilibrium, drivers' belief is consistent with realized supply and demand distributions.

At the end of a period, an unmatched driver of firm f in location i makes a search decision after observing supply shocks $\{\epsilon^j\}_{\forall j}$ by choosing the location with maximum value:

$$j^* = \arg\max_j \{V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \rho_f \underbrace{(V^j_{ft+\chi^{ij}_t} - \min_l \{V^l_{ft+\chi^{il}_t}\})}_{\Delta^j_{ft}} \mathbb{1}_{\chi^{ij}_t=1} + \epsilon^j_f\} \tag{6.12}$$

where c^{ij}_t is the cost of travelling from i to j calculated as $c^{ij}_t = 0.75 * distance^{ij}_t$. The cost per mile is set to be 0.75 dollars. $V^j_{ft+\chi^{ij}_t}$ is driver's ex-ante value of searching location j in period $t + \chi^{ij}_t$ before the matching process in period $t + \chi^{ij}_t$. The number of periods travelling from i to j at t is χ^{ij}_t which is time cost compared to c^{ij}_t. Drivers are assumed not to pick up passengers along his way to the search location. When the driver chooses j which is far from i, he has to account for the loss of not searching for passengers until the next χ^{ij}_t periods. This time costs plays two important roles in the model. First, it contributes to mismatches across locations due to mobility. For example, a location i has many vacant cars at the end of period t, while in $t + 1$ there are many passengers in j far from i. This may result in excess supply in locations near j but excess demand in j at $t + 1$. Second, I can study benefits of traffic improvement by changing χ^{ij}_t. Both c^{ij}_t and χ^{ij}_t are allowed to vary over periods t and route i, j, but common to Uber and taxis. These two variables can be directly calculated from the data. At last, the parameters ρ_f measures extra benefits of searching locations that are close to current location i, $\chi^{ij}_t = 1$. The difference of $V^j_{ft+\chi^{ij}_t}$ minus the minimal

[27]This number is obtained by counting the number of outcomes of putting N balls in I different urns $C^{I-1}_{N+I-1} = \frac{(N+I-1)!}{(I-1)!(N)!}$ allowing for empty urns.

values guarantee non-negative benefit. The ex-ante value is defined as:

$$V^j_{ft} = \phi^j_{ft}\big(\sum_l \tilde{a}^{jl}_{ft}(p^{jl}_{ft} - c^{jl}_t + V^l_{ft+\chi^{jl}_t})\big) + \\ (1-\phi^j_{ft})\mathbb{E}_\epsilon\big[\max_l\{V^l_{ft+\chi^{jl}_t} - c^{jl}_t + \rho_f \Delta^j_{ft}\mathbb{1}_{\chi^{ij}_t=1} + \epsilon^l_f\}\big]. \quad (6.13)$$

In equation (6.13), ϕ^j_{ft} denotes the matching probability of drivers, $\phi^j_{ft} = m^j_{ft}/v^j_{ft}$. Conditional on being matched, the expected profit is obtained by averaging over all possible destinations l with weights $\tilde{a}^{jl}_{ft}$. Recall that $\tilde{a}^{jl}_{ft}$ measures firm specific destination distribution obtained in (6.11). The profit conditional on completing trip jl includes the fare of the trip, cost of travelling, and continuation value in location l after dropoff in $t + \chi^{jl}_t$ period.

The second part of (6.13) is the continuation value of not being matched in j. The interpretation of each variable is the same as (6.12). Since drivers do not observe realized supply shock ϵ's until the end of period, the continuation value takes expectation over all possible supply shocks. I assume the supply shock ϵ_f to follows i.i.d T1EV distribution with scale parameter σ_f for each firm f such that the continuation value of unmatched has an explicit form:

$$\mathbb{E}_\varepsilon \max_l\{V^l_{ft+\chi^{jl}_t} - c^{jl}_t + \rho_f \Delta^j_{ft}\mathbb{1}_{\chi^{ij}_t=1} + \epsilon^l\} = \sigma \log\big(\Sigma_l \exp((V^l_{ft+\chi^{jl}_t} - c^{jl}_t + \rho_f \Delta^l_{ft}\mathbb{1}_{\chi^{il}_t=1})/\sigma_f)\big) \quad (6.14)$$

Given the feature of supply shock's distribution, I can calculate deterministic transition probability of unemployed drivers such that the probability of an unemployed driver of firm f in location i searching j in the next period is:

$$\pi^{ij}_{ft} = \frac{\exp((V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \rho_f \Delta^j_{ft}\mathbb{1}_{\chi^{ij}_t=1})/\sigma_f)}{\Sigma_l \exp((V^l_{ft+\chi^{il}_t} - c^{il}_t + \rho_f \Delta^l_{ft}\mathbb{1}_{\chi^{il}_t=1})/\sigma_f)} \quad (6.15)$$

The scale parameter σ_f controls for incentives of drivers searching certain locations captured by continuation values other than shocks. For example, large σ_f implies that drivers' search decisions are largely driven by random supply shocks which leads to an even allocation of drivers' searches across locations.

Recall the transition of employed cars following the dropoffs distribution of passengers $\{\tilde{a}^{ij}_{ft}\}$ calculated in (6.11). Combining the transition of employed cars $\tilde{\boldsymbol{A}}_{\boldsymbol{f}}$ and

the policy function of unemployed cars $\mathbf{\Pi_f}$ of equation (6.15) gives the law of motion for state transition. The state includes the status of all in-transit cars. The state at the beginning of period t is a collection of $\{S_t^i\}_{\forall i}$ where S_t^i is a collection of $\{\tilde{v}_{ft,k}^i\}_{f,k}$ with $\tilde{v}_{ft,k}^i$ indicating the number of cars for firm f that will arrive at location i in the next k periods. When $k = 1$, it implies that the supply at period t satisfies $v_{ft}^i = \tilde{v}_{ft,k=1}^i$. At the end of each period, the transition of employed and unemployed cars update the state such that:

$$\tilde{v}_{ft+1,k}^i = \tilde{v}_{ft,k+1}^i + \Sigma_j m_{ft}^j \tilde{a}_{ft}^{ji} \mathbb{1}_{\chi_t^{ji}=k} + \Sigma_j (v_{ft}^j - m_{ft}^j) \pi_{ft}^{ji} \mathbb{1}_{\chi_t^{ji}=k}, \forall f, i, k \tag{6.16}$$

To interpret (6.16), at beginning of period $t+1$, the number of firm f drivers that will arrive at location i in k periods is composed of three parts: (1) those who will arrive at i in $k+1$ periods at the beginning of period t; (2) those who pickup passengers at time t and will arrive at i in k periods; (3) those unemployed drivers of period t who decide to search location i next but will arrive in k periods. In the next section, I discuss the matching function applied to calculate matches in the demand and supply decisions.

6.3 Matching Function

During the matching process in each period within a location, I use an explicit functional form to predict the matching outcomes. In the works of Buchholz (2016) and Frechette et al.(2016) studying NYC taxi industry, they all assume a matching process with friction for taxis within a location. Frechette et al.(2016) simulate the process of taxis searching over grids within a location for passengers. Buchholz (2016) assumes an urn-ball random matching process and a corresponding explicit functional form can be derived by Burdett, Shi and Wright (2001). I use the same functional form as Buchholz (2016) and Burdett, Shi and Wright (2001) with a modification to reflect heterogeneity in frictions across locations. This matching process is only applied to taxis within locations outside two airports. For taxi in airports and Uber in all locations, I assume perfect matching within the location. I will discuss both functions below.

First, I introduce the random matching of taxis in locations other than airports. Given taxis' demand u_{yt}^i and supply v_{yt}^i in location i at period t, I assume that passengers randomly visit the taxis and of those visiting the same car only one can be successfully matched. Other unmatched passengers will leave with the subway. I do

not distinguish where passengers and cars are located within the market such that all cars are identical to the passengers. That means a passenger has equal probability visiting any car. Each car receives a passenger's visit with probability $1/v^i_{yt}$. Therefore, the probability of a taxi not receiving a visit is $(1-1/v^i_{yt})^{u^i_{yt}}$. The probability of a taxi being matched is $1-(1-1/v^i_{yt})^{u^i_{yt}}$. However, in locations with larger size of area, it is different to know the exact location of cars. Thus, I add an location specific parameter γ_i such that the probability of a car being matched becomes $1-(1-1/(\gamma_i v^i_{yt}))^{u^i_{yt}}$. Higher value of γ_i will decrease the probability of being matched. To be specific, I define $\gamma_i = \gamma_1\{i \in CentralManhattan\} + \gamma_2\{i \in OuterBorough\}$. Since taxis within i have the same probability of being matched, and therefore the number of matches is:

$$\begin{aligned} m(u^i_{yt}, v^i_{yt}) &= v^i_{yt}\Big(1-(1-\frac{1}{\gamma_i v^i_{yt}})^{u^i_{yt}}\Big) \\ &\approx v^i_{yt}(1-\exp(-\frac{u^i_{yt}}{\gamma_i v^i_{yt}})) \end{aligned} \tag{6.17}$$

Function (6.17) itself allows matching friction due to coordination failures such that there is possibility that some cars receive no visits and some passengers are not matched.

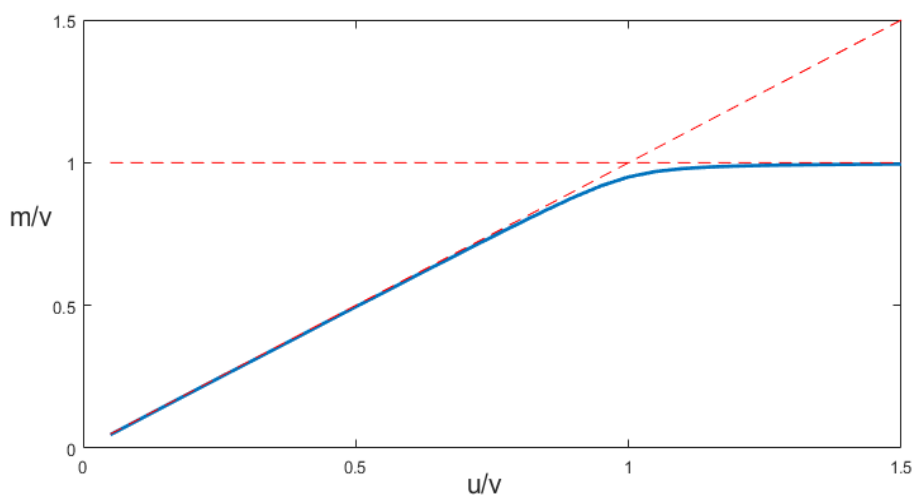


Figure 4: Hyperbolic function for perfect matching

As for Uber and aiports, the matching between drivers and passengers is assumed to be frictionless. Uber use mobile technology to assign passengers and drivers into a one-to-one pair without coordination failure as above. As for airport, drivers are

waiting in a queue to match passengers one by one without coordination failure as well. However, I do not model the queueing process as Buchholz (2016) for simplicity. The explicit functional form of perfect matching is $m^i_{xt} = \min\{u^i_{xt}, v^i_{xt}\}$. However, one drawback of this function is that it is not always invertible given any two variables for the third variable. For example, when $m^i_{xt} = v^i_{xt}$ the solution of u^i_{xt} is not unique. Instead, I use a hyperbolic function to approximate perfect matching as shown in figure 4. When the demand-to-supply ratio is greater than 1, the matching probability of drivers approaches 1. When the ratio is less than 1, the matching probability approaches 45 degree line that is $m^i_{xt} \approx u^i_{xt}$. The explicit matching function form of figure 4 is obtained by solving m^i_{xt} from (with small value of ϵ):

$$(\frac{m^i_{xt}}{v^i_{xt}} - 1)(\frac{m^i_{xt}}{v^i_{xt}} - \frac{u^i_{xt}}{v^i_{xt}}) = \epsilon \tag{6.18}$$

Even when $m^i_{xt} = v^i_{xt}$, it can returns an u^i_{xt} though its value depends on the curvature rather than any economic assumption. In the next section, I will discuss equilibrium of the model in details.

6.4 Equilibrium

I use the equilibrium concept of OE such that unemployed drivers make search decision based on his own state and knowledge about state evolution of the market. The key information about the market state is the distribution of supply and destination distribution of in-transit cars. Driver's own state is denoted as s_t which includes his location at time t. The state of the city at time t is denoted as collection $\{\mathcal{S}^i_t\}_{i\in I}$. For any i, $\mathcal{S}^i_t$ includes information about arrival of cars in next K periods, hence collection $\{\tilde{v}^i_{ft,k}\}_{f,k\in K}$. In OE, drivers make optimal search decisions according to $\{s_t, \{\mathcal{S}^i_t\}_{\forall i,t}\}$. Drivers' belief on the evolution of market state (i.e. supply distribution) is consistent with the realized state in equilibrium[28]. Given the deterministic evolution of supplies, equilibrium demand can be simply calculated for each market due to the static discrete choice assumption. The definition of equilibrium is summarized as follows:

Definition Equilibrium is a sequence of supply $\{v^i_{ft}\}$, beliefs of state transition $\{\hat{v}^i_{ft,k}\}$, policy function of unemployed cars $\{\pi^{ij}_{ft}\}$, transition of employed cars $\{\tilde{a}^{ij}_{ft}\}$ for $\forall i, t, f$

[28] Another way to understand the OE in this model is that instead of knowing evolution of supply distribution, drivers know the evolution of ex-ante search values $\{V^i_{ft}\}_{\forall i,t}$. Knowing supply distribution or search values are interchangeable given one step calculation of (6.8).

and given initial distribution of supply $\{v^i_{ft=1}\}$ such that:

1. At the beginning of each period t, in any location i, passengers make discrete choice between firms based on (6.1)-(6.8). Market demand is calculated from (6.10).

2. Matches are made randomly between supply and demand for each firm, location and time. Within location, the matching process follows (6.17) for taxis and (6.18) for Uber and taxi in airport.

3. Transition of employed cars follows $\{\tilde{a}^{ij}_{ft}\}$ obtained by Bayes' rule (6.11).

4. At the end of each period, unemployed drivers follow policy function π^{ij}_{ft} calculated in (6.15) based on beliefs of state transition $\{\hat{v}^i_{ft+1,k}\}$.

5. Realized state transition is obtained by combining both employed $\{\tilde{a}^{ij}_{ft}\}$ and unemployed cars $\{\pi^{ij}_{ft}\}$. State of next period is updated to $\tilde{v}^i_{ft+1,k}$ by (6.16).

6. At the beginning of next period, both employed and unemployed cars arrive and form the new supply $\tilde{v}^i_{ft,k=1} = v^i_{ft}$.

7. Drivers' belief is consistent such that $\hat{v}^i_{ft+1,k} = \tilde{v}^i_{ft+1,k}$ for all i, t, f, k

My model's equilibrium is quite similar to Buchholz(2016) and still satisfies finite horizon and finite action-space for existence of equilibrium. In the next section, I will explain estimation process. The main idea of estimating this model is to solve unobserved equilibrium demand and supply such that the equilibrium model-generated matches fit the pickups observed in the data for all well defined markets $\{i, t\}_{i\in I, t\in T}$.

7 Estimation

In this section, I will discuss the estimation process of my model in detail. The key feature of this estimation is that supply and demand level in any given market are not directly observed in the data. Instead, the data only has observation of pickups, which are the outcome of underlining matching process given supply and demand. Thus, estimation of the model is searching for parameter values, equilibrium demand and supply that generate outcomes fitting the data. The demand side parameters include

mean utility $\{\delta^i_{ft}\}_{\forall f,t,i}$, price coefficients $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, and substitution parameter β. The supply side parameters include supply shock parameter σ_f, bonus of searching nearby locations ρ_f. Finally, there are parameters defining the matching function of taxis $\{\gamma_1, \gamma_2\}$. Given the values of these parameters, I can solve equilibrium supply $\boldsymbol{v_f}$ and demand $\boldsymbol{u_f}$, $f = 1, 2$. Then parameters in $\boldsymbol{\delta}$ can be estimated using linear regression.

This process of estimating these parameters can be summarized as figure 5. Given demand parameter values and market size, market demand can be calculated. Then, given supply side parameters values and drivers' optimal decisions I can calculate supply using backward induction due to the finite time of the model. Finally, the matching function can predict pickups given supply and demand above. The model is estimated once the model predicted pickups equal pickups in the data and model generated transitions of employed taxis follow the same dropoff distribution of taxis in the data.

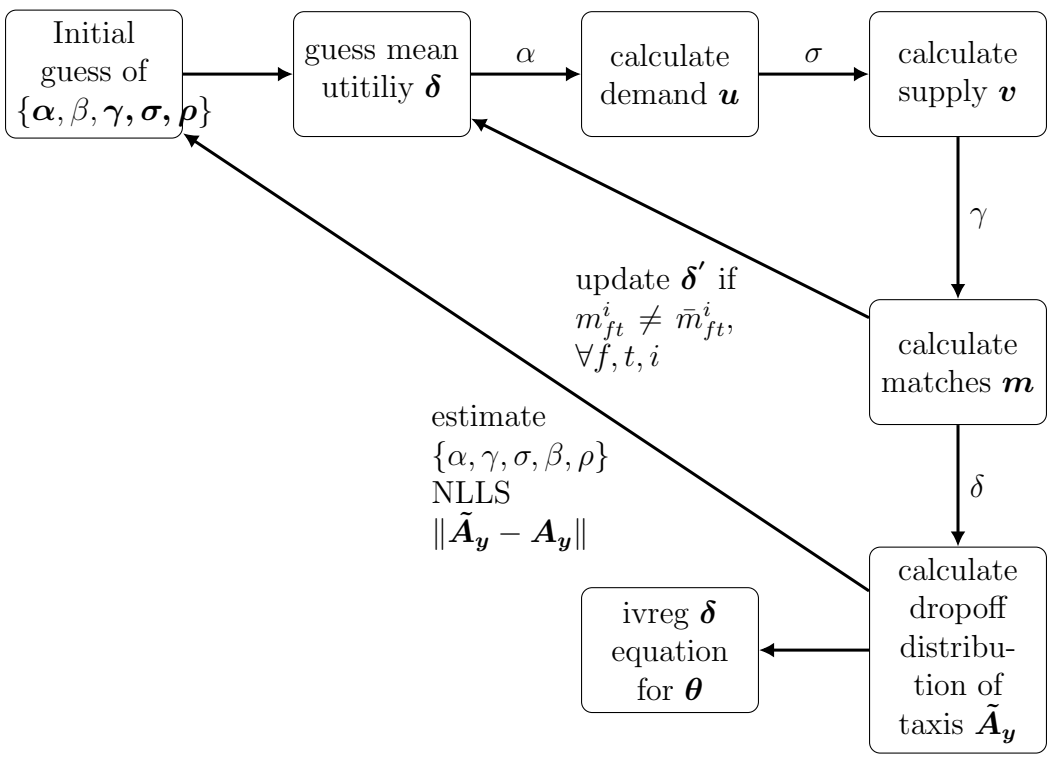


Figure 5: overview of the estimation process

In general, the process is similar to estimating a random coefficient discrete choice model but with more complexities. Given a set of nonlinear parameter values $\{\alpha, \gamma, \sigma, \beta, \rho\}$, I solve the fixed point for mean utilities $\boldsymbol{\delta}$ such that the model predicted shares of

pickups equal shares in the data[29]. Given $\boldsymbol{\delta}(\alpha, \gamma, \sigma, \beta, \rho; \bar{m})$, I estimate nonlinear parameters using nonlinear least square estimation to diminish the difference between model predicted dropoffs of taxis to the dropoffs in data. Parameters within $\boldsymbol{\delta}$ are estimated using linear IV regression because of endogeneity of supply $\boldsymbol{v}$ and demand $\boldsymbol{u}$ in demand equation. The instrument variable I use for supply and demand is dropoffs of cars from the same and opponent firm in the market and market size. Recall that supply of a market comprises arrivals of employed and unemployed drivers. Demand shock influences arrival of unemployed drivers but employed drivers drop passengers off according to exogenous destination of passengers regardless of the demand in that destination. In the following subsections, I first discuss NLLS estimation as step 1 followed by IV regression as step 2.

7.1 Pre-estimation Discussion

7.1.1 Number of Taxi and Uber Drivers

In order to predict the supply distribution over locations and time, I have to fix the total number of cars both for taxis and Uber. The supplies are intractable if drivers can enter or exit in any location and at any time. Even with fixed number of cars, I need to assume an initial supply distribution of cars at period 1. I assume the fixed number of taxis in my model as 13,587 which is the total number of medallions[30]. As for Uber, though there are around 26,000 Uber affiliated cars, most of them only work part-time or in weekends. In Chen et al. (2017), they have detailed data about Uber drivers' working hours. They define types of driving schedules of Uber drivers into evening, morning, late-night, weekend and infrequent categories and find a transition matrix of Uber drivers among these types of schedule across weeks as cited in figure 6. For example, in the first row, 11.0% of evening driver in this week will switch to morning drivers in the following week. In my paper, working patterns of individual drivers is irrelevant since I do not distinguish identities. I obtain the fixed number of daytime drivers by calculating the stationary distribution of the markov chain in figure 6. In the stationary distribution, there are 10% morning drivers which is equivalent to

[29]The definition of nonlinear parameters here follows Nevo (2000). He denotes parameters outside mean utilities as nonlinear parameters and parameters within mean utilities as linear parameters.

[30]Buchholz (2016) uses the number of medallions as total taxis in his model. Frechette et al.(2016) also find that almost 80% of minifleets and 70% of self-owned taxi medallions are active during day shift hours.

around 3,000 Uber cars in NYC. Thus, I assume the number of active Uber cars in my sample hours from 6 am to 4 pm to be 3,000.

Type of Driver	% of Active Drivers	t \ $t+1$	Eve	Morn	Late	Wknd	Infr
Evening Driver	16.1	Eve	38.3	11.0	9.3	21.0	20.4
Morning Driver	5.9	Morn	12.1	13.4	8.3	18.1	48.1
Late-Night Driver	6.1	Late	17.2	16.8	19.4	17.3	29.2
Weekend & Evening Driver	19.3	Wknd	8.7	5.4	3.2	70.6	12.1
Infrequent Driver	52.6	Infr	7.1	11.1	4.9	9.8	67.0

Figure 6: Uber drivers' working schedule table cited from Chen et. al.(2017)

7.1.2 Passengers' Destinations

As mentioned in the sample construction, in the demand side I need to know the exogenous destinations of the travelling population composed of taxi & Uber passengers and subway passengers paying full fare. It is important to note that travellers with subway pass or taking other transportation tools (i.e. bike, bus, and walking) are not considered. There is no data available on the city transportation pattern. Thus, I use taxis' pickup-dropoff pattern of 2010 as proxy for the population travelling pattern in my model. The two implicit assumptions here are that: first, the pattern doesn't change due to Uber's entry; second, subway passengers paying full fare follow the same travelling pattern as taxi passengers. Thus, the destination distribution of population, denoted as $\{A_t^i\}$ in the model, conditional on markets of origin can be nonparametrically calculated using taxis trip records in 2010.

7.2 Two-Step Estimation

7.2.1 Step 1: Estimating Nonlinear Parameters

The first step is to estimate $\{\boldsymbol{\delta}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \boldsymbol{\sigma}, \boldsymbol{\rho}\}$. The estimation process can be further broken into two procedures. First, given the parameter values of $\{\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \boldsymbol{\sigma}, \boldsymbol{\rho}\}$, I need to solve equilibrium demand as function of mean utilities $\boldsymbol{\delta}$ and supply such that model generates the same pickups as observed in the data. Second, in the outer loop of first procedure, I update $\{\boldsymbol{\delta}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \boldsymbol{\sigma}, \boldsymbol{\rho}\}$ to minimize deviation of model predicted distribution of dropoffs for matched taxis to the dropoff distribution in data. The

first procedure is similar to Buchholz (2016) with an extra calculation due to the introduction of discrete choice demand to the model. In Buchholz (2016), he searches for equilibrium demand $\boldsymbol{u}$ which generates pickups in the sample. Instead of solving for demand $\boldsymbol{u}$, I am solving for fixed mean utilities $\boldsymbol{\delta}$ which have one-to-one mapping to demand according to BLP contraction mapping. Solving mean utilities is necessary to generate destinations of matched passengers for each firm $\tilde{\boldsymbol{A}}_{\boldsymbol{f}}$ which forms objective function of the second procedure.

In details, the first procedure is as follows. Given parameter values $\{\alpha, \sigma, \gamma, \rho, \beta\}$, I make initial guess of all mean utilities $\{\delta^i_{ft}\}_{\forall f,t,i}$ which are obtained by using pickups as demands to calculate market share and applying BLP contraction mapping to back out mean utilities. Given the initial guess mean utilities and demands, I need to figure out corresponding equilibrium supplies $\{v^i_{ft}\}_{\forall f,t,i}$. Instead of an arbitrary guess of supplies, I use backward induction to calculate my starting point for equilibrium supply. In details, from the last period T, given the demand distribution and zero search values $V^i_{ft} = 0, \forall t > T$, I calculate continuation values $\{V^i_{fT}\}$ for an arbitrary supply distribution $\{v^i_{fT}\}$[31]. For period $T-1$, given the search values $\{V^i_{fT}\}$ and demands $\{u^i_{fT-1}\}$, I can calculate search values $\{V^i_{fT-1}\}$ and updated supplies of next period $\{v^i_{fT}\}$ for an arbitrary supply distribution $\{v^i_{fT-1}\}$. Repeat this process backward until period $t = 1$. The supply distribution in period 1 is assumed to be proportional to pickups distribution of period 1. This process returns my starting values of supplies $\boldsymbol{v}$. I iterate the process of updating supplies and search values from $t = 1$ forward to $t = T$ until supplies and search values stop updating such that equilibrium supplies are obtained. The whole procedure is summarized in algorithm 1 below. I find starting point of supplies for iteration in step 2-14. Step 15-17 is the iteration for equilibrium supplies[32]. This procedure generates mapping from mean utilities to equilibrium demands $\boldsymbol{u} = u(\boldsymbol{\delta})$ and supplies denoted as $\boldsymbol{v} = \Gamma(\boldsymbol{\delta})$. The pickups generated by model are $\boldsymbol{m} = m(u(\boldsymbol{\delta}), \Gamma(\boldsymbol{\delta}))$. In order to fit pickups to the monthly average pickups $\bar{\boldsymbol{m}}$, I need to update mean utilities in the outer loop of the procedure above. The outer loop is summarized in algorithm 2.

The second procedure follows the first one to estimate $\{\alpha, \sigma, \gamma, \rho, \beta\}$ given the

[31]The day ends for taxi drivers in period T due to shifts. However, Uber drivers have no shifts. Uber drivers may continue to work after period T with positive search values. In this estimation, I assume search values are equal over locations for $t > T$ and normalized to 0. Normalization won't affect transition probabilities since constants are cancelled out by equation 6.15.

[32]This process does not satisfy contraction mapping. In order to find the fixed point solution, I applies iterative method of average damping.

Algorithm 1 Solve Equilibrium Supply

1: Set parameter values for $\{\sigma, \alpha, \gamma, \rho, \beta\}$ and $\{\delta^i_{ft}\}_{\forall f,i,t}$
2: Guess supply $\{v^i_{fT}\}_{f,i}$, calculate $\{u^i_{fT}\}_{f,i}$, $\{V^i_{fT}\}_{f,i}$
3: for $\tau = T-1$ to 1
4: Guess $\{v^i_{f\tau}\}_{f,i}$, reset state of in-transit cars $\{\tilde{v}^i_{ft,k}\}_{t>\tau} = 0$
5: for $t = \tau$ to $t = T$
6: Compute market share $\{s^i_{ft}\}$ and demand $\{u^i_{ft}\}$
7: Compute matches $m^i_{ft} = m(u^i_{ft}, v^i_{ft})$
8: Compute transition of employed cars $\{\tilde{a}^i_{ft}\}$
9: Compute transition policy of unemployed cars $\{\pi^i_{ft}\}$
10: Update $\{\tilde{v}^i_{ft+1,k}\}$ based on $\{\pi^i_{ft}\}, \{\tilde{a}^i_{ft}\}$
11: Update value $\{V^i_{ft}\}$ of current period
12: Update supply of next period $v^i_{ft+1} = \tilde{v}^i_{ft+1,k=1}$
13: end
14:end
15:Fix $\tau = 1$
16: Iterate step 5 to 13
17: Stop until step 11 and 12 won't update under certain tolerance level.

Algorithm 2 Solve Fixed Points of Mean Utility

1: Guess demand $\{u^i_{ft}\}_0$ based on observed pickups $\{\bar{m}^i_{ft}\}$
2: Calculate market share s^i_{ft} and guess initial $\{\delta^i_{ft}\}_0$
3: Iterate BLP contraction mapping to solve for $\{\delta^i_{ft}\}_1$ to match market shares
4: Plug $\{\delta^i_{ft}\}_1$ into algorithm 1 to solve for $\{v^i_{ft}\}$
5: Invert matching function with $\{v^i_{ft}, \bar{m}^i_{ft}\}$ for $\{u^i_{ft}\}$
6: a: $v^i_{ft} > \bar{m}^i_{ft}$, update u^i_{ft}
7: b: $v^i_{ft} \leq \bar{m}^i_{ft}$, don't update u^i_{ft}
8: Given updated $\{u^i_{ft}\}_1$, solve BLP contraction mapping for $\{\delta^i_{ft}\}_2$ and $\{\tilde{A}^i_{yt}\}$
9: a: $\Sigma_f u^i_{ft} < \lambda^i_t$, update δ^i_{ft}
10: b: $\Sigma_f u^i_{ft} \geq \lambda^i_t$, set $u^i_{xt} = u^i_{ft}\bar{m}^i_{xt}/\Sigma_f u^i_{ft}$ and $u^i_{yt} = \lambda^i_t - u^i_{xt}$, update δ^i_{ft}
11: Repeat step 4 to 10
12: Until $|\boldsymbol{\delta}^{k+1} - \boldsymbol{\delta}^k| < \epsilon$
13: Report $\{\tilde{A}^i_{yt}\}$, transition of employed taxis

transition probability of matched cars $\{\tilde{A}^i_{yt}\}$. For each of the 60 time periods, the transition probabilities form a 40 by 40 matrix. Instead of matching the probabilities point-to-point to probabilities in the data, I aggregate the pickups and dropoffs over locations and periods and recalculate the transition probabilities in larger areas over 20 half hours. This process applies to both data and model generated transitions. Then I calculate the sum of squared deviations between model generated transition of taxis to the observed transition of taxis in data as my nonlinear least squares objective function. Estimators are defined as 7.1 below:

$$\{\hat{\alpha}, \hat{\sigma}, \hat{\gamma}, \hat{\rho}, \hat{\beta}\} = \underset{\alpha,\sigma,\gamma,\rho,\beta}{\arg\min} \Sigma_{t,i} \big(\bar{m}^i_{yt}(A^{ij}_{yt} - \tilde{A}^{ij}_{yt}(\alpha, \sigma, \gamma, \rho, \beta))\big)^2 \tag{7.1}$$

7.2.2 Step 2: Estimating Demand Parameters

Given the estimates of $\{\Theta, \boldsymbol{m}, \boldsymbol{\delta}, \boldsymbol{u}, \boldsymbol{v}\}$ with $\Theta = \{\alpha, \sigma, \gamma, \rho, \beta\}$ in step 1. I can calculate variables in mean utility equation that are not directly obtained in data such as market demand and supply in equation (6.2). In the specification of mean utility, the coefficient on demand measures direct network effect and coefficient on supply measures indirect network effect. An OLS regression of equation (6.2) suffers endogeneity problem since demand and supply are all correlated with unobserved demand shock ξ^i_{ft}. For supply v^i_{ft}, I use arrival of employed cars of firm f at the beginning of period as instrument. The argument is that these cars visit location i because of their passengers' destination. It is reasonable to assume that demand shock of current period is not correlated with the destination of passengers picked up from other locations in previous periods. One exception to this assumption could be that some passengers visit location i to pick someone up and leave i immediately. This instrument is correlated with supply as it constitute supply together with arrival of unemployed cars. To solve to endogenous problem of demand, I use two instruments, market size λ^i_t and arrival of opponent's cars. Market size is exogenous and correlated with demand as in (6.10). Arrival of opponent's cars are uncorrelated with demand shocks following the same argument above. Moreover, it also correlates with demand by affecting the choice probability of passengers.

7.3 Identification

The parameters are identified by variation of pickups, the travelling pattern of population and the pattern of taxi passengers' dropoffs over locations and time in the data. Given a set of nonlinear parameter values, mean utilities $\{\delta^i_{ft}\}$ are identified by the variation of pickups across firms and markets. The mapping from mean utilities to pickups follows algorithm 1&2 in which I firstly map mean utilities to demands and corresponding dynamic supply distributions followed by calculating matches given supply and demand. Consider two identical markets (i.e. same market size, travelling pattern and prices) with different pickups $\bar{m}^1 > \bar{m}^2$. In a static model, market 1 with higher pickups implies a higher demand and supply than market 2 and therefore $\delta^1 > \delta^2$. However, in a dynamic model, the corresponding supplies for given mean utilities are more complicated than in static model. In the dynamic model, supply may not fully respond to demand variation across locations due to mobility restriction of cars conditional on their locations in the previous period. However, given the one-to-one mapping from inter-period demands to dynamic supplies, the dynamic pickup patterns in the data helps to identify mean utilities.

Identification of price coefficient α comes from variation of population travelling patterns over markets. For example, two markets with same mean utility $\delta^i_{ft} = \delta^j_{ft}$ and market size λ^i_t but with different destination distributions of passengers $A^i_{ft} \neq A^j_{ft}$ will have different unconditional(on destination) demand $u^i_{ft} \neq u^j_{ft}$. The demand level relative to subway ridership also helps identifying price coefficient. As for supply shock parameter σ, it controls for transition of unemployed drivers. Given the search values over locations $V^i_{ft}, \forall i$, high σ implies equal probability of search in each location i.

Identification of matching function parameters γ is not quite intuitive. They are crucial to connect the mapping from δ to pickups. Different γ do not affect equilibrium supply as much as equilibrium demand. The reason is that drivers' matching probability does only depend on successful pickups rather than potential demand. Given the pickups generated by model equal those of data in estimation, matching efficiency does not change probability much. However, given supply level fixed, inefficient matching of γ affects estimated mean utilities δ. For example, given fixed number of pickups and supply, a large γ (less efficient) will generate high potential demand and corresponding high δ. The mean utilities further affects destinations of passengers in equation (6.11) and drivers' profits. Since ex-ante search values V^i_{ft} depend not only on matching probability but also on profits, equilibrium supply also reacts to change of γ. Thus, in order

to identify the $\{\delta^i_{ft}\}$ with restriction to γ, I use the dropoffs of taxis in the data such that model predicted dropoffs of taxis match the data. The reason is that different magnitudes of δ, which is shifted by γ, not only affect market shares relative to outside option but also affects distribution of firm specific dropoffs as in (6.11). High δ could dominate heterogeneous price effects on utilities over different routes and making the distribution of taxis' dropoffs close to destinations of market population[33]. Hence, I use taxis' dropoffs to identify the matching function parameters.

Finally, linear parameters in δ are estimated using instrument variables. Since both demand and supply are endogenous and correlate with demand shocks, I need to find instrument for demand and supply. I choose arrival of employed cars from the same and opponent firm as instruments which are correlated with supply and demand but uncorrelated with demand shocks. The estimation result is in next section.

8 Results

8.1 Estimates outside mean utilities

The estimation results are listed in table 4. The estimates of price coefficients for different trip distance are $\{\hat{\alpha}_k\}_{k=1,2,3,4}$ as specified in (6.3). The price coefficient for trip distance less than 3 miles is -0.81 and becomes less sensitive for long distance trip. The estimate $\hat{\alpha}_4 = -0.26$ is for trips between JFK and Manhattan which charges flat rate for taxi passengers. The estimate of another demand parameter $\beta \in [0, 1]$ in nested logit demand defined in (6.5) is equal to 0.38. When $\beta \to 1$, demand shocks for taxi and Uber are highly correlated and when $\beta \to 0$ they are independent as in a simple logit model.

There are two sets of parameters from the supply side. First, the estimates of supply shocks' scales $\sigma_f, f = y, x$ for taxi and Uber are 7.67 and 12.65. These two parameters affect the transition probability of unemployed cars as in (6.15). Larger σ_f implies less effect of profit differences (without accounting for supply shocks) across locations on transition probability π^{ij}_{ft}. The other way around, a smaller σ_f will enlarge the difference in profits among locations such that drivers have higher likelihood to

[33]For a given market i, t, the mean utility δ^i_{ft} is common for all destinations j. It shifts the conditional shares on routes uniformly. In the extreme case of taxis' mean utility large enough, destination of taxis' passengers is exactly same to population's.

search high profit area. The estimates means that taxi drivers have higher incentive to search for locations with higher search values than Uber drivers. Furthermore, controlling for profit differences across locations, these two estimates implies a higher chance that taxi drivers will overcrowd high profit areas and leave low profit areas undersupplied. In other word, σ_f enhance the role of profit gap caused by regulated price on matching frictions across locations. Second set of supply parameters are ρ_f which measures the incentive of unmatched drivers to search locations nearby in the next period. This extra bonus from current location i to search location j is measured proportionally to $V^j_{ft+\chi^{ij}_t} - \min_l\{V^l_{ft+\chi^{il}_t}\}$ if $\chi^{ij}_t = 1$ (see 6.12). This proportion for taxi is 0.38 and for Uber is 0.27. Taxi's higher $\hat{\rho}_y$ means taxi drivers have slightly higher incentive to search locations nearby than directly visiting a location far away.

Table 4 also reports the estimates for random matching function 6.17. The $\gamma_i = \gamma_1\{i \in CentralManhattan\} + \gamma_2\{i \in OuterBorough\}$ measures within market matching efficiency for taxis in non-airport locations. I distinguish the efficiency in central Manhattan and outer Borough (Northern Manhattan, Brooklyn and Queens). Higher value of γ means less efficient matching within the market. For instance, given a fixed number of supply and demand, higher γ generate less successful matches. The estimates for this parameter in Manhattan area is 1.11 in comparison to other locations which is 3.67. Given these estimates, it indicates that the within market matching is less efficient in Outer Borough. It is reasonable considering on the large area sizes of defined markets in Queens and Brooklyn.

Finally, there are 4800 mean utilities $\{\delta^i_{ft}\}_{\forall f,t,i}$ to estimate and the statistics for taxi or Uber is listed in the bottom of table 4. The mean of taxis' $\{\delta^i_{yt}\}$ is 1.21 with maximum value at 4.90 and minimum value at -1.79. Uber's mean utilities $\{\delta^i_{xt}\}$ are less than taxis' after controlling for prices. It is because Uber's supply and demand are still far less than taxis. Market share of taxis are still much higher than that of Uber. Even controlling for prices, higher market share of taxis corresponds to higher mean utility level. One novel interpretation is that the number of taxi riders and drivers have strong network effects on passengers' choice between taxi and Uber. By the specification of δ^i_{ft} in (6.2), taxis' high mean utility could result from network effects.

Along with the parameter estimates, the statistics of three key variables solved in equilibrium including demand, supply, and search values are shown in table 5 for taxi and Uber respectively. The mean of taxis' demand in 2400 markets(location-period) is 118.13 with maximum demand of 531.61 and minimum of 1.66. In comparison to taxis,

Table 4: Estimates of nonlinear parameters

panel 1: nonlinear parameter		*estiamtes*
demand side parameters		
$\hat{\alpha}_1$		-0.81
$\hat{\alpha}_2$		-0.58
$\hat{\alpha}_3$		-0.41
$\hat{\alpha}_4$		-0.26
$\hat{\beta}$		0.38
supply side parameters		
$\hat{\sigma}_y$		7.67
$\hat{\sigma}_x$		12.65
$\hat{\rho}_y$		0.38
$\hat{\rho}_x$		0.27
matching function		
$\hat{\gamma}_1$		1.11
$\hat{\gamma}_2$		3.67
mean utilities	*mean*	*min/max*
$\hat{\delta}^i_{yt}$	1.21	-1.79/4.90
$\hat{\delta}^i_{xt}$	0.30	-1.62/4.02

Uber's potential demand is less with 20.92 on average. The average supply of taxis for each market is 174.21 ranging from 3.02 to 1779.3. The average supply of Uber is 34.46. Finally, the ex-ante search values across markets for taxis have the maximum value of \$194.33 dollars at the beginning of the day (6 a.m.). Uber's maximum search values is \$273.27 dollars at the beginning of the day and the average value is \$128.75 dollars. In general, Uber drivers have higher expected profit than taxis. This high profitability of Uber could be the result of surge pricing, matching efficiency and competition within firm. Specifically, the minimum fare \$7 and surge multiplier make the expected profit of Uber conditional on being matched higher than taxis. Technology makes the matching probability of Uber higher than taxi controlling for supply and demand in a market. In addition, the number of active Uber cars are much less than taxis such that within firm cannibalization is smaller than taxis.

Table 5 only shows the statistics of demand, supply and search values over both locations and periods. Given the dynamic feature of this model, search values follow a decreasing trend over t which is not revealed in simple statistics. Thus, I draw four figures to demonstrate four key values by firm, by location over time. The first one

Table 5: Statistics in equilibrium

demand	*mean*	*min/max*
u^i_{yt}	118.13	1.66/531.61
u^i_{xt}	20.92	1.06/136.65
supply		
v^i_{yt}	174.21	3.02/1779.3
v^i_{xt}	34.46	4.45/227.7
$ search values		
V^i_{yt}	98.25	1.10/194.33
V^i_{xt}	128.75	4.87/273.27

is figure 7 which shows the values of V^i_{ft} over t. Each line represents a location i and there are 40 lines. Given the finite periods of this dynamic game, the values decrease over time. Uber's values are generally larger than taxis' due to its higher conditional expected profit and probability of being matched.

By 6.13, the search value is sum of two parts: expected profit of being matched and expected continuation value of not being matched. I separately show each of the two conditional expected values in figure 8, 9. There are two findings by comparing these two figures with figure 5. First, by comparing figure 5 and 8, I find that conditional expected profits on being matched are more dispersed than search values V^i_{ft}. It implies that matching probability plays a role to shrink the gaps of profitability among locations and therefore leads to much more close search values of them. For instance, drivers are self-motivated to search locations with high conditional expected profits which decrease the matching probability of that location such that search value of a location with high conditional expected profit approaches search value of a low profitable location. Second, by comparing figure 8 and 9, I find that expected values conditional on being matched are higher than continuation values conditional on being unmatched. It means that drivers are always willing to be matched than to be unmatched in any market. This finding is not redundant given the assumption of my model since it does not hold in certain instances. For example, my model assumes that drivers arriving at a market at the beginning of the period must attend the matching process of that market. If a driver drops off a passenger in a market where he has a higher probability to pick up a new passenger travelling to a destination with a quite low search value, he would rather to be unmatched and make search decisions.

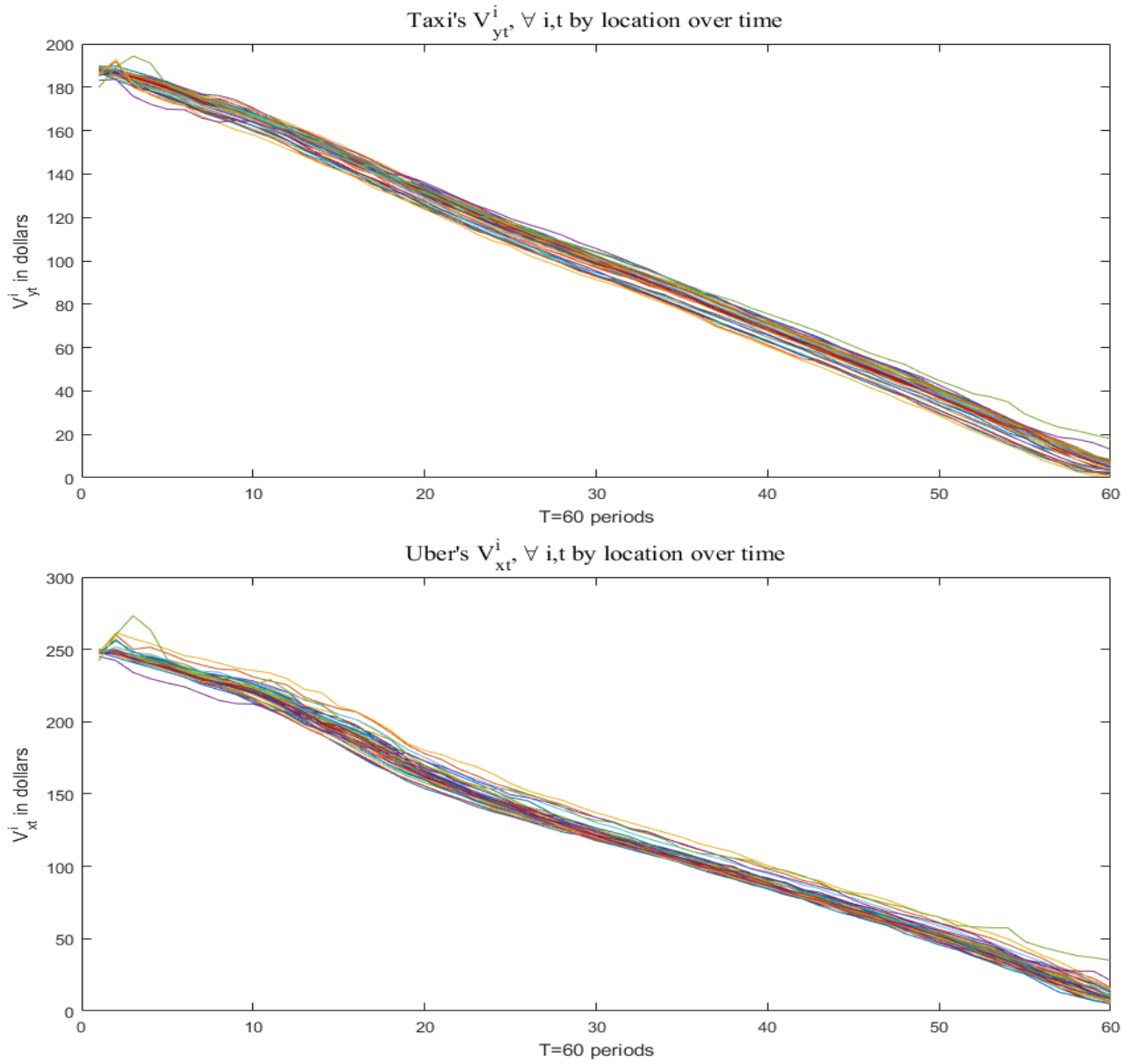


Figure 7: Dynamics of search values before matching process

In order to compare the differences of $V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \rho_f \Delta^j_{ft} \mathbb{1}_{\chi^{ij}_t=1}$ across destination j for a given location i, I draw figure 10. The differences of $V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \rho_f \Delta^j_{ft} \mathbb{1}_{\chi^{ij}_t=1}$ directly determine the search choices of drivers based on equation (6.15). For any market i, t, unmatched driver needs to choose a location to search based on optimization decision 6.12. I calculate two times standard deviation, $2 * std_j\{V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \rho_f \Delta^j_{ft} \mathbb{1}_{\chi^{ij}_t=1}\}$, to represent the differences in incentives of drivers to search among locations. A small value of this difference means drivers have equal incentive to search any of the 40 locations. In figure 10, the horizontal axis is time periods and vertical axis is the difference. Each line represents a location i over t and the variation of continuation values a driver face if unmatched. The figure shows that when unmatched drivers make search decisions, they face a heterogeneous values among search options. The 2*standard deviation of these values range from \$6 to \$14 in $t = 1$ among 40 markets for taxis. For Uber, the standard deviation changes largely over time and is greater than taxis. To interpret the comparison, after controlling for scale parameter of supply shock, σ_f, Uber drivers have more incentive than taxis to search certain locations.

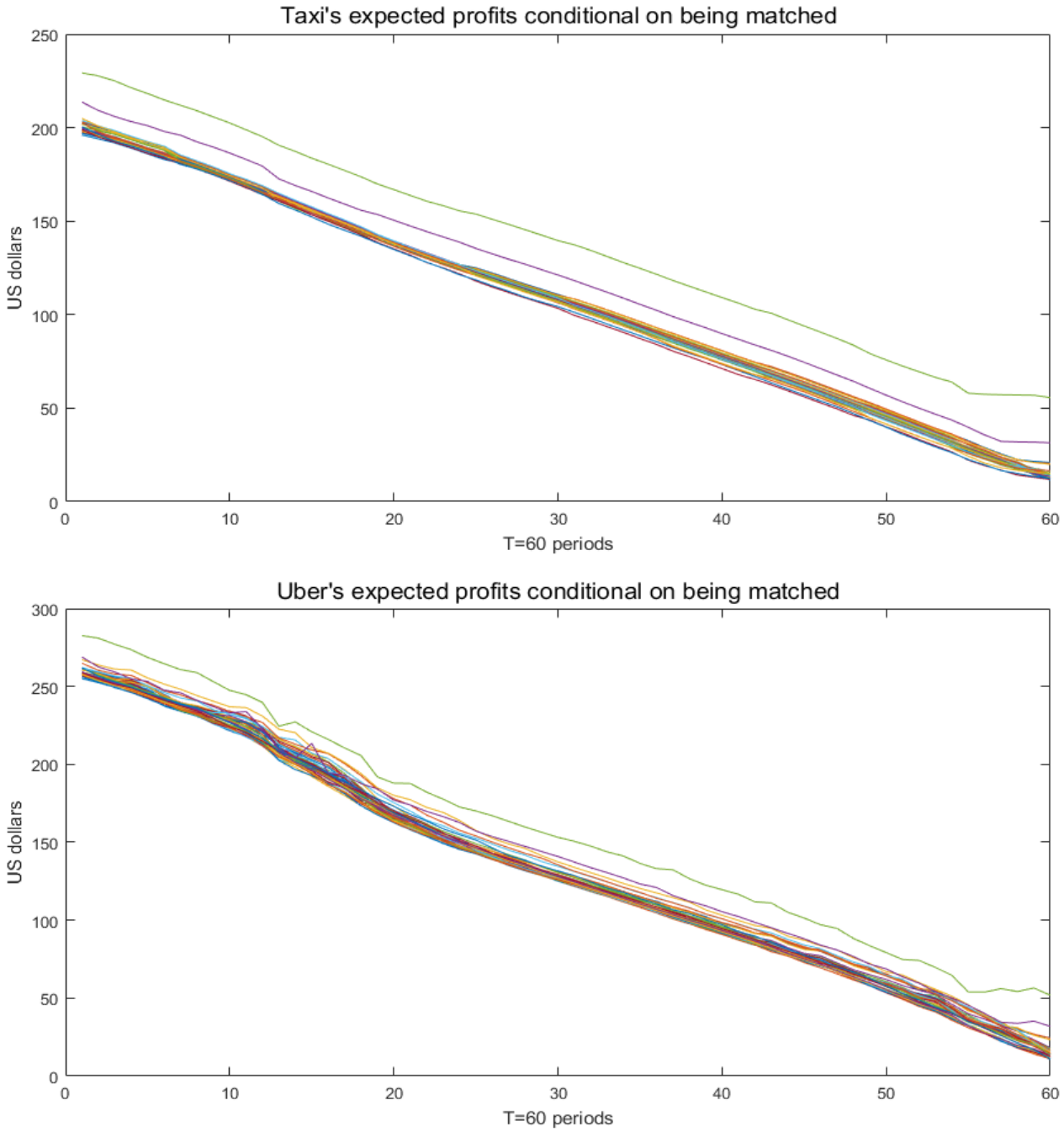


Figure 8: Dynamics of expected profits conditional on being matched

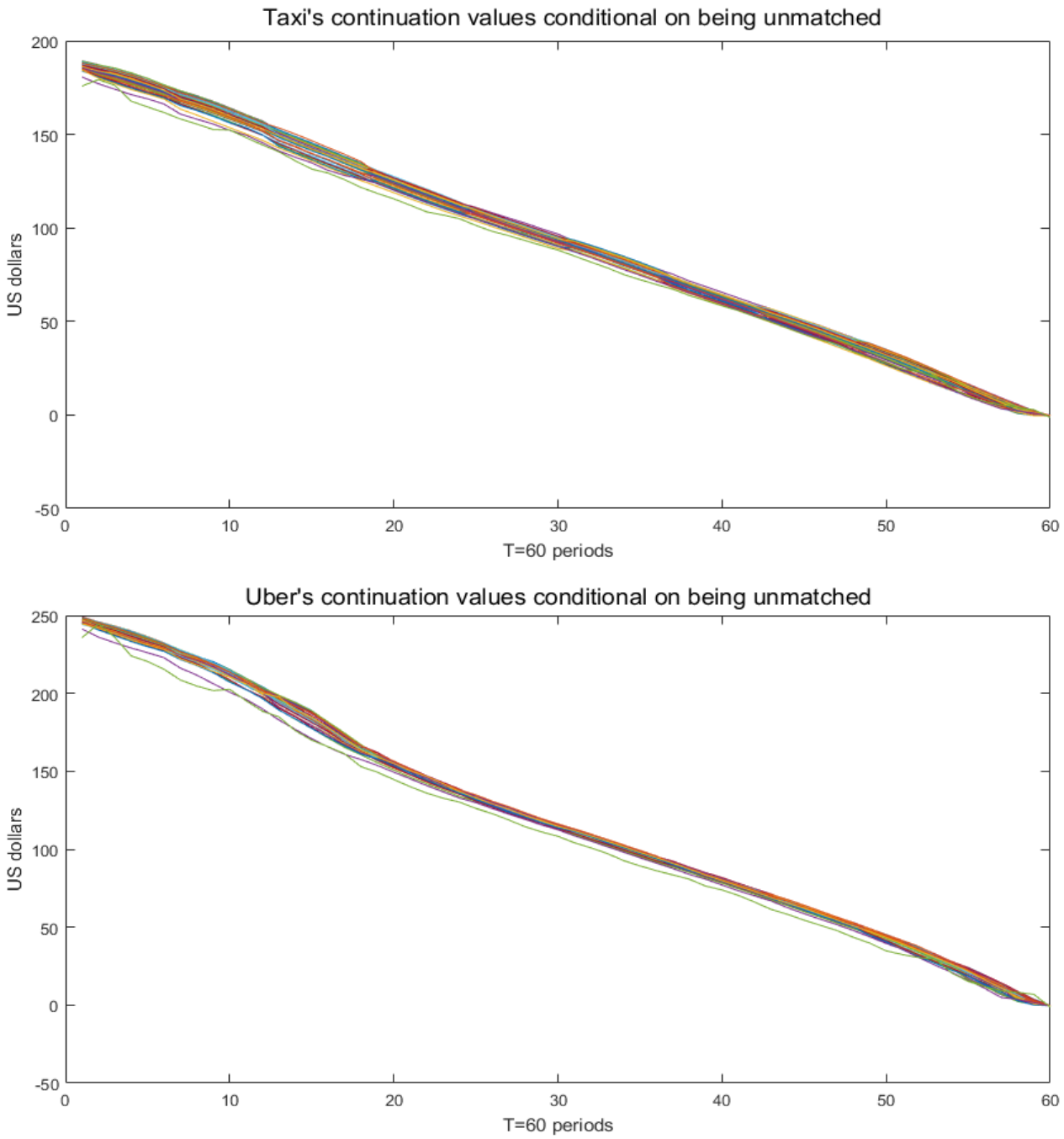


Figure 9: Dynamics of expected continuation values conditional on being unmatched

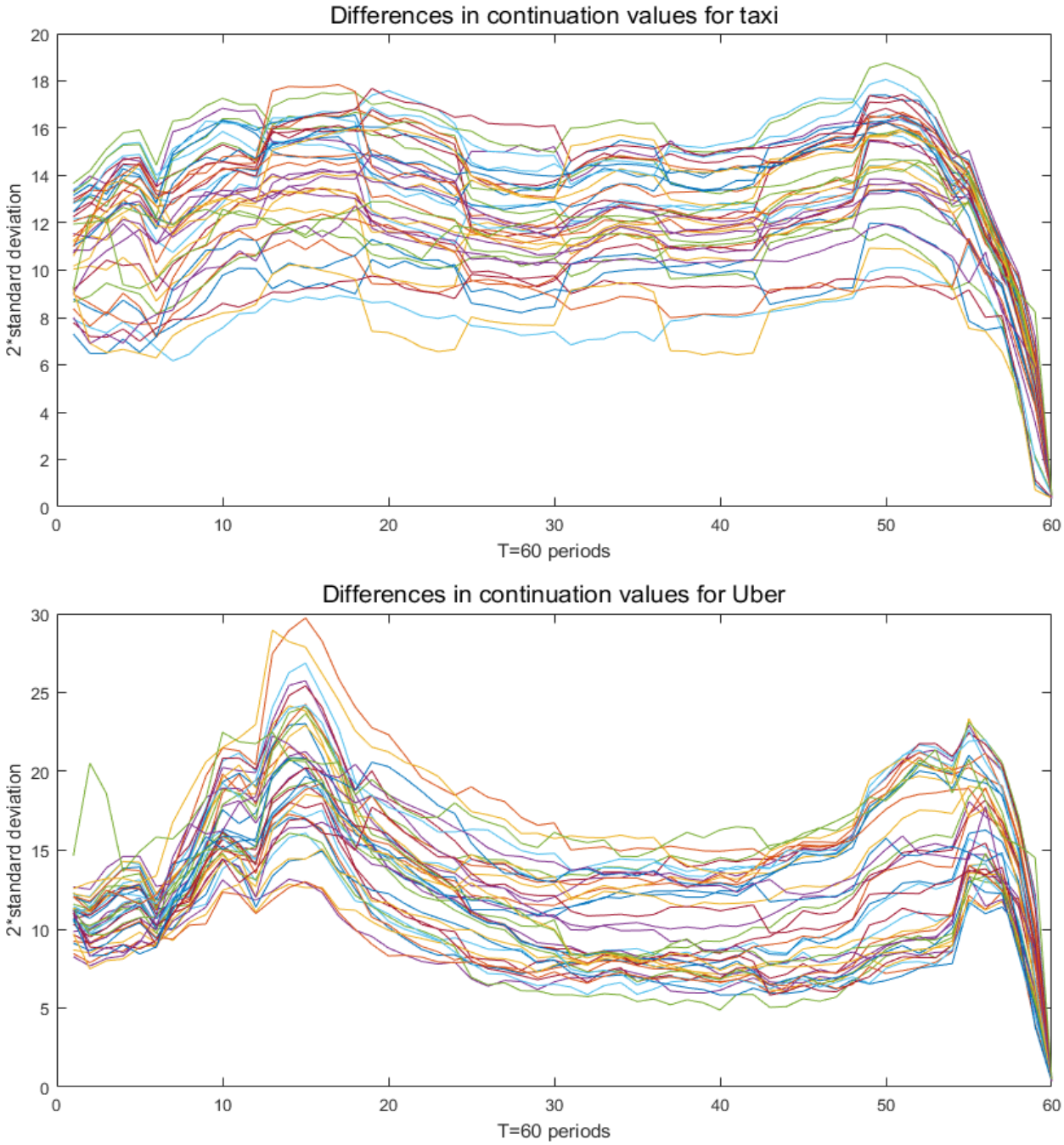


Figure 10: Dynamics of expected continuation values conditional on being unmatched

8.2 Estimates in mean utilities

Given previous estimation results, I run a linear regression of mean utilities $\{\delta^i_{ft}\}$ on variables in equation (6.2). The coefficients of interest are θ_1, θ_2. The supply coefficient measures indirect network effects from the other side of the market. It is a net effect of supply on utility of choosing product including, but not limited to, impact via matching probability and waiting time. Likewise, the coefficient on demand captures net effect of demand on utility. One possible channel is that the demand decreases the chance of being matched and therefore negatively affect utility. The other way around, it can also positively affect utility of choices via consumers learning from each other, herding and culture. The cause of this joint effects, as discussed in section 6.1, is approximation of matching probability and log linear assumption of ex-post utility, see 6.2.

I use market size, arrival of drivers from the same and opponent firm as instruments for $\ln u, \ln v$. The OLS and 2SLS regression results are shown in table 6. In the regression, I add interaction term of Uber dummy with logarithm of demand and supply levels. The estimates show that both positive effects of demand and supply on utility. Moreover, the effects are larger for utility of choosing taxis than for Uber. The fixed effect of Uber on mean utility is positive which equals 1.21. The positive sign of supply coefficient means that higher supply level increases the choice probability of taxi(or Uber) and thus demand level. Since demand will positively affect supply as implicitly imposed in supply side structure, these two effects form the positive feedback loop between drivers and passengers[34]. It is a little surprised that the coefficient on demand is also positive. Demand is expected to negatively affect matching probability after controlling for supply level. One way to explain this negative sign is that there exists strong positive direct network effect among passengers which offset the negative effect on probability. For example, if my neighbor, colleague or friend uses taxi (or Uber), I would also like to choose taxi (or Uber). To highlight early, the coefficients on demand and supply in the mean utility are important for forming feedback loop in two sided market. In the counterfactual of this model, I compare the scenarios with or without network effect by allowing the mean utility to react to changes of demand and supply or not. This exercise will generate two different equilibria for understanding the consequences of ignoring network effects.

[34]In the supply side, I do not parameterize and estimate the effects of demand on drivers' supply decisions. Instead, the positive effect is "imposed" because matching probability of drivers increases in demand AND conditional expected profit is higher than expected continuation value as discussed.

Table 6: Linear regression of mean utility

Dependent variable δ^i_{ft}		
	OLS	*2SLSIV*
$\ln v$	0.054	0.496
	(0.019)	(0.07)
$\ln v \times d_x$	-0.053	-0.224
	(0.039)	(0.11)
$\ln u$	0.53	0.249
	(0.021)	(0.06)
$\ln u \times d_x$	-0.014	-0.086
	(0.038)	(0.08)
Uber dummy d_x	0.19	1.21
	(0.054)	(0.18)
constant	-2.226	-3.71
location fixed effects	YES	YES
time fixed effects	YES	YES

8.3 Benchmark Welfare

In this section, I will discuss and analyze the matching efficiency of this industry given the model backed out demand, supply and observed matches, prices and so on. The key factors that I am interested in are the two types of mismatches in the model, within location friction and cross location friction. Within location friction is partly reflected by the mismatches between drivers and passengers of the same firm in the matching process. Since the matching process within a market is assumed, this type of friction is mainly driven by the functional form (6.17) though introducing γ adds flexibility to the function. Cross location mismatches means in the same period some locations have more drivers than demand whereas some locations have more demand than supply before the matching process within locations. Excess supply of a location can be counted as $\max\{v^i_{ft} - u^i_{ft}, 0\}$ and likewise excess demand is counted as $\{u_{ft} - v_{ft}, 0\}$. These expressions does not account for the mismatches due to random matching within market such that I can distinguish these two types of frictions. In period t, the city level aggregate demand is $\Sigma_i u^i_{ft}$ and aggregate supply is $\Sigma_i v^i_{ft}$. The maximum aggregate matches that can be made without type 1 friction are $\Sigma_i \min\{u^i_{ft}, v^i_{ft}\}$. Given the aggregate demand and supply level fixed, the efficient matches should be $\min\{\Sigma_i u^i_{ft}, \Sigma_i v^i_{ft}\}$ from

the city aggregate perspective[35]. The difference is:

$$
\begin{aligned}
&\min\{\Sigma_i u^i_{ft}, \Sigma_i v^i_{ft}\} - \Sigma_i \min\{u^i_{ft}, v^i_{ft}\} \\
&= \min\{\Sigma_i u^i_{ft} - \Sigma_i \min\{u^i_{ft}, v^i_{ft}\}, \Sigma_i v^i_{ft} - \Sigma_i \min\{u^i_{ft}, v^i_{ft}\}\} \\
&= \min\{\underbrace{\Sigma_i \max\{0, u^i_{ft} - v^i_{ft}\}}_{\text{aggregate excess demand}}, \underbrace{\Sigma_i \max\{v^i_{ft} - u^i_{ft}, 0\}}_{\text{aggregate excess supply}}\}
\end{aligned}
\tag{8.1}
$$

Expression 8.1 counts the minimum of aggregate excess supply and demand. Unlike assumed within location friction, this friction is mainly driven by the endogenous supply and demand decisions of drivers and passengers. To be a good measure of friction, we should be able to compare efficiencies of two scenarios by their index values. There are three limitations for the validity of (8.1) as a measure of friction. First of all, it only counts the static mismatches in a given period. An less efficient matches of current period could make better matches in the next period considering the mobility of drivers across locations. Second, it only counts the number. But passengers and trips are not identical. I improves this measure by weighting number by its dollar value. Third, even without previous two limitation, this index is invalid if aggregate excess supply is larger than aggregate excess demand. The reason is that aggregate demand is not a fixed number as aggregate supply does. A large value of this index could still end up with a large number of aggregate matches.

The welfare statistics are listed in table 7. The first panel displays type 1 mismatches both in terms of counts and dollar values. For example, there are totally 95,547 within location mismatches for all 40 locations from 6 a.m.- 4 p.m. of a weekday. These mismatches could generate 1.3 million for taxis if within location matching is perfect. In other words, without coordination failure within market, drivers could make 50% more profits. For now, taxi drivers' daily total profit is 2.5 million. The within location mismatches for Uber is negligible. Though I assume perfect matching for Uber within market, the hyperbolic function still generate slight friction than perfect matching. The cross location mismatches as measured by 8.1 are shown in second panel. There are totally 14,738 type 2 mismatches for taxis which could make $203,530 trip fares. Uber has a far less cross location mismatches which is 1,850 in counts and

[35]Lagos (2000) treats this expression as efficient aggregate matches and aggregate matching function generating less matches has friction. However, in his paper, the demand is exogenous and there is no demand-supply feedback loop.

\$34,450 in money value. The revenue made by all taxi drivers in a day shift is 2.5 million and Uber drivers make \$ 779,380. Consumer welfare is evaluated by inclusive value of logarithm utility. In the counterfactuals, I will compute the compensating variation to compare consumer welfare changes.

Table 7: Baseline Welfare Statistics

<u>within-location mismatches</u>		
$\Sigma_{i,t} \min\{u^i_{yt}, v^i_{yt}\} - \bar{m}^i_{yt}$		95,547
		\$1,286,400
$\Sigma_{i,t} \min\{u^i_{xt}, v^i_{xt}\} - \bar{m}^i_{xt}$		635
		\$10,517
<u>cross-location mismatches</u>		
$\Sigma_t \min\{\Sigma_i \max\{u^i_{yt} - v^i_{yt}, 0\}, \Sigma_i \max\{v^i_{yt} - u^i_{yt}, 0\}\}$		14,738
		\$203,530
$\Sigma_t \min\{\Sigma_i \max\{u^i_{xt} - v^i_{xt}, 0\}, \Sigma_i \max\{v^i_{xt} - u^i_{xt}, 0\}\}$		1,850
		\$34,450
<u>Profits and welfare</u>		
taxi profit		\$ 2,510,400
Uber profit		\$ 779,380
consumer welfare		505,210
<u>matches: data v.s. model generated</u>	data	model
$\Sigma_{i,t} m^i_{yt}$	173,490	173,230
$\Sigma_{i,t} m^i_{xt}$	48,897	47,738

The dynamics of frictions over periods are provided in figure 11. It shows how the aggregate frictions and matches over locations for any given period evolve. Taxi pickups increase sharply after the first hour in the morning and decrease until $t = 30$ at 11 a.m.. Uber's aggregate pickups are more flat than taxis which only increases slightly during rush hours of morning. Notably, the cross location mismatches reach the daily highest level during the morning rush hours. In other words, during rush hours, taxi drivers are more likely to overcrowd some areas and leave others undersupplied. At last, I show several figures for selected locations about their demand, supply and pickups of taxi and Uber over time, see figure 12 to 15. One interesting finding by comparing Queens or Brooklyn with central Manhattan areas is that there are excess taxis' supply in Manhattan almost all the time during the daytime. In comparison, Queens and Brooklyn are more likely to have excess demand. Especially during the morning hours from 7 am to 9 am. It is harder for passengers to hail a taxi. Another

finding is that Uber serves outer boroughs much in comparison to taxis. The number of Uber drivers in selected location of Brooklyn is close to that in Times Square. In next section, I will simulate several counterfactuals to learn how they affect the matching efficiency of this market.

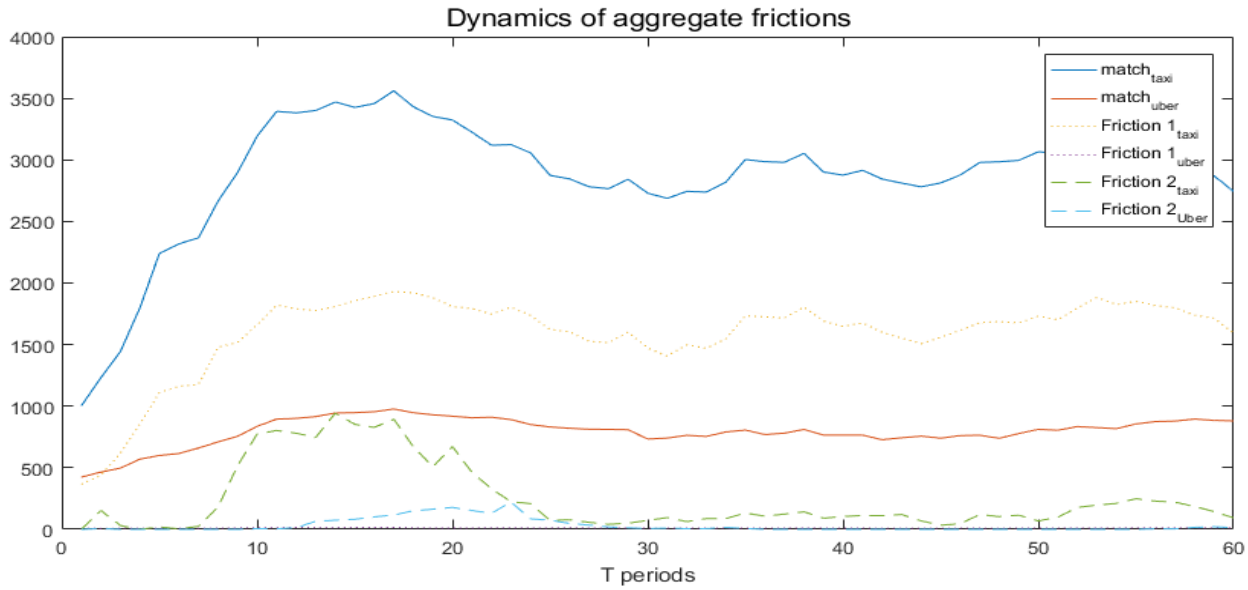


Figure 11: Aggregate frictions and matches of taxis& Uber over time

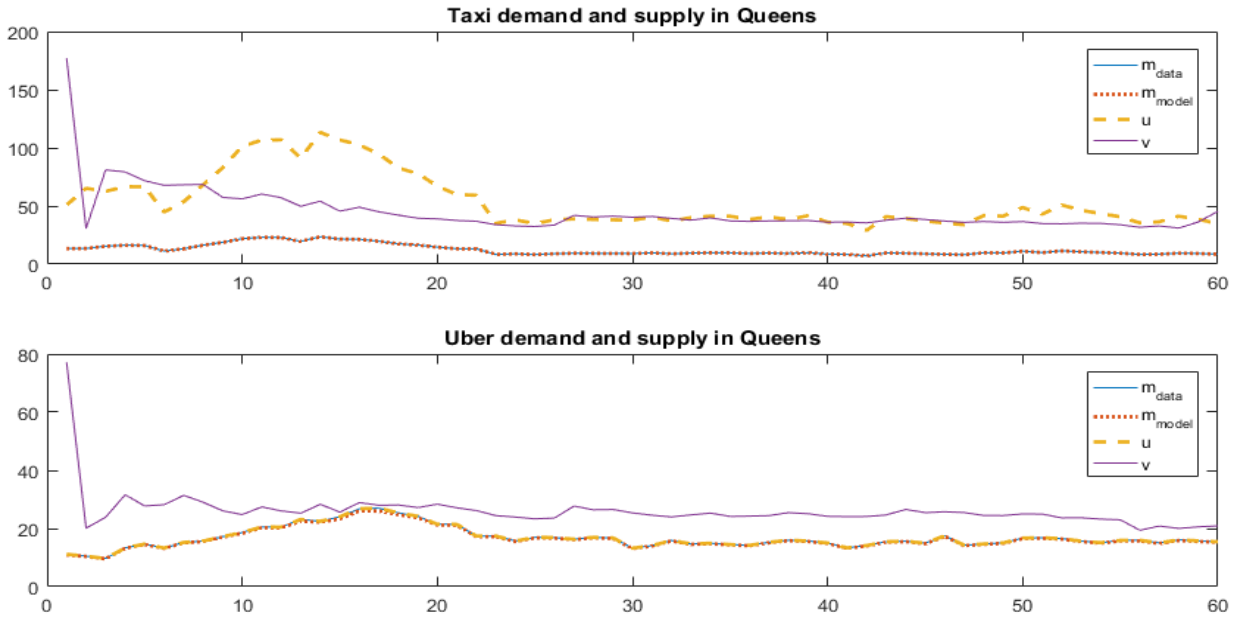


Figure 12: Demand, supply and matches in a location of Queens

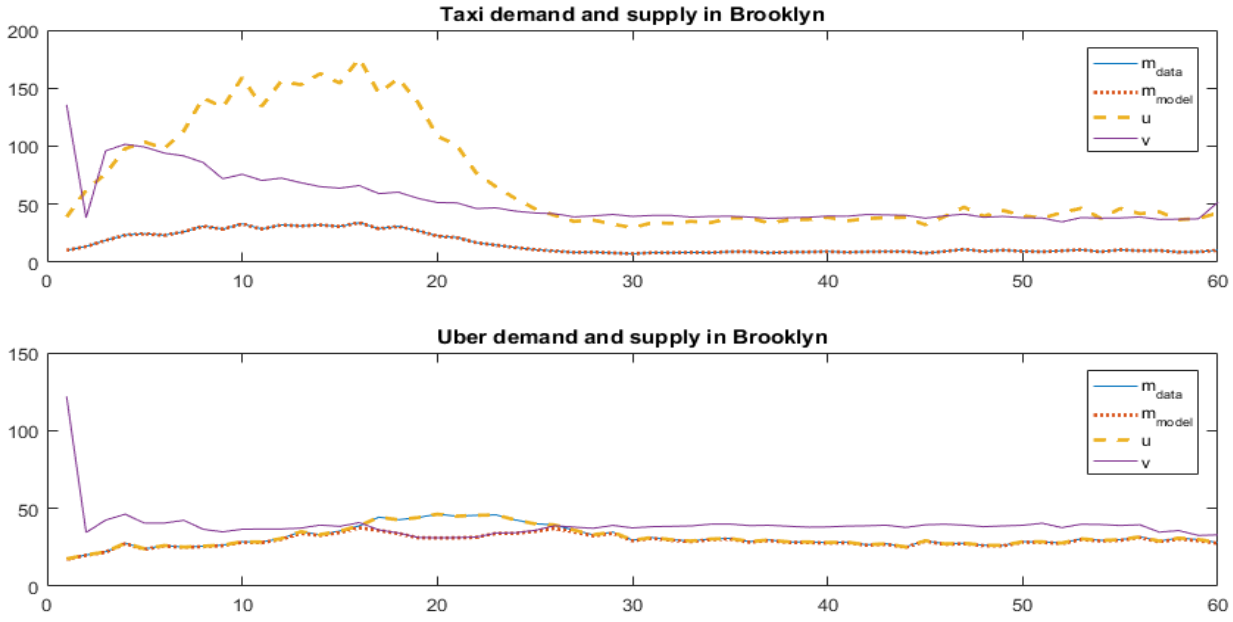


Figure 13: Demand, supply and matches in a location of Brooklyn

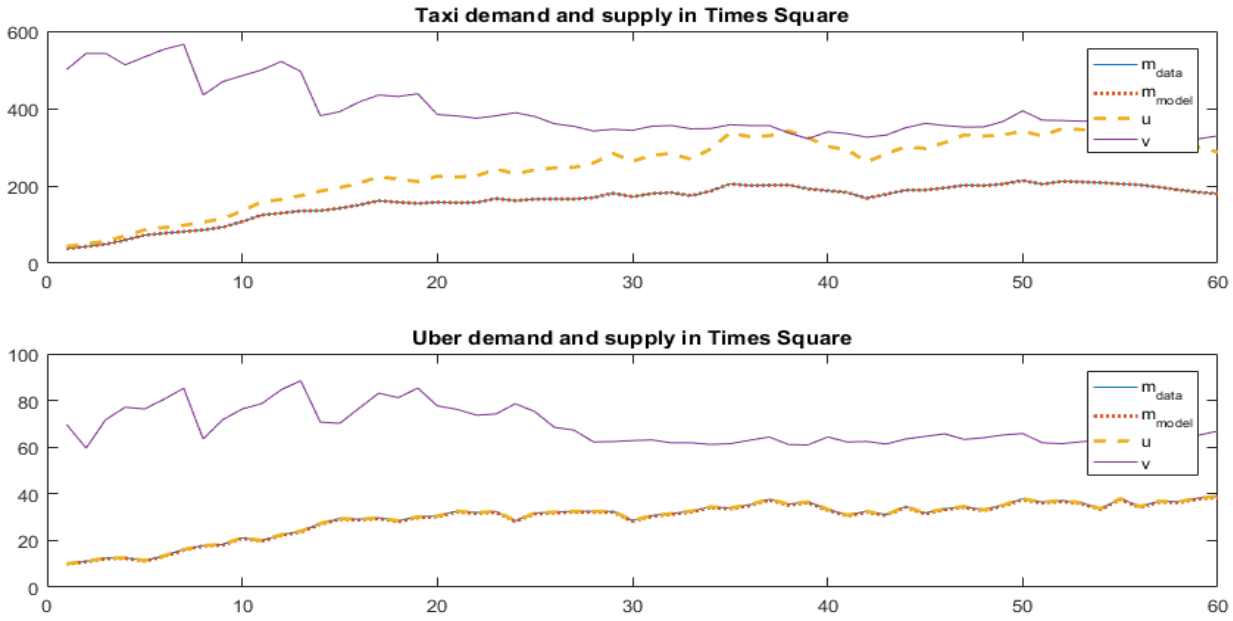


Figure 14: Demand, supply and matches in Times Square

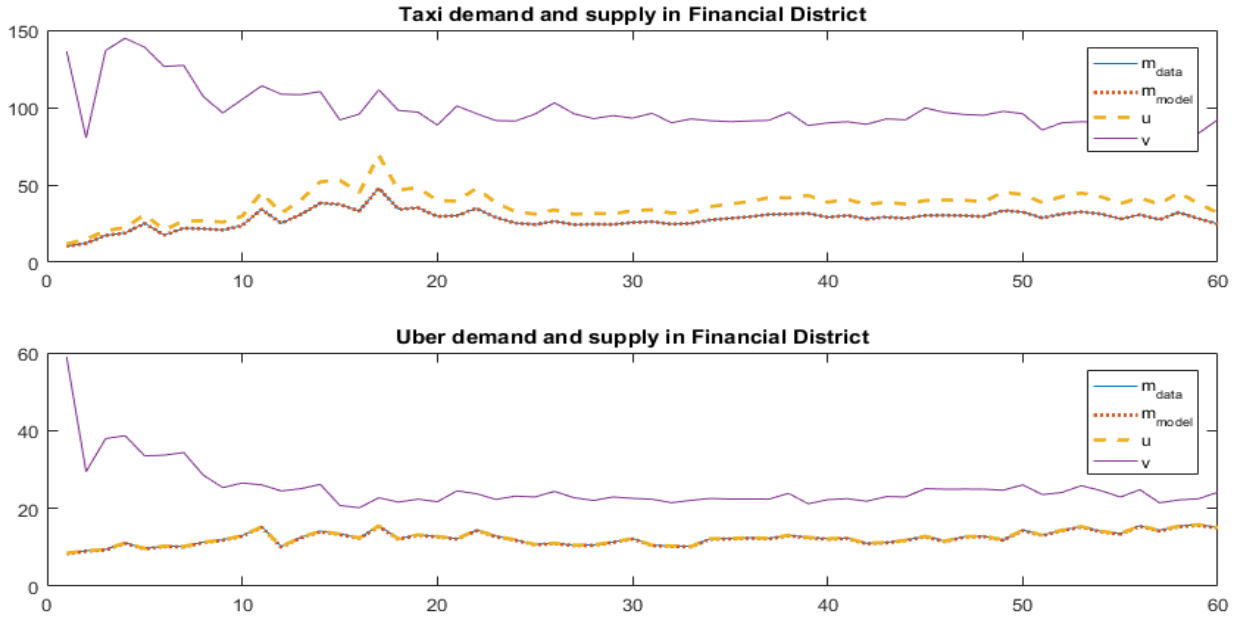


Figure 15: Demand, supply and matches in Financial District

9 Counterfactuals

In the previous sections, I estimate a dynamic search and matching model of taxi & Uber drivers and passengers. The results show that: 1, there exists feedback loop (indirect network effect) between demand and supply in a market (platform) and direct network effect within the same side; 2, when drivers make search decisions, they face a very heterogeneous search values among locations; 3, drivers are more likely to oversupply high profitable location and leaving other locations undersupplied such that cross location mismatches exist; 4, the low matching probability due to oversupply in high profitable location reduces the ex-ante value of that location such that V_{ft}^{i} are close for different locations i. This paper focuses on the "mis-allocation" of drivers, therefore the type 2 friction, because type 1 friction is driven by the random matching assumption. In order to understand what factors and to what extent they affect the matching efficiency and social welfare, I simulate two counterfactuals at the beginning of this section, followed by counterfactuals of regulatory policy. First, I analyze how Uber surge pricing affects efficiency of matching. Second, I study to what extent traffic condition matters for matching. As for the policies, I study the government's proposal to cap Uber in 2015 and its effects. I also simulate the congestion pricing policy that the NYC government is thinking of to reduce the congestion in Manhattan.

Before discussing each simulation, the simulation process that applies to all counterfactual scenarios is listed below. The different between with or without network effect is whether or not to update mean utility $\boldsymbol{\delta_f}$ for the new equilibrium demand and supply. Not updating mean utility means that passengers will not respond to the change of demand and supply level so that feedback loop between two sides is shut down. This case is similar to Buchholz 2016 without taking into account of network effects. With network effect, the mean utilities will also adjust to any change of supply levels and so does demand. Furthermore, the supply will respond to the change of demand and so forth until new equilibrium is reached. One note is that in the new equilibrium, I assume the demand shocks $\hat{\boldsymbol{\xi}}_f$ backed out from linear regression are fixed.

Simulation Algorithm Without Network Effects

1: Fix parameter values as estimates $\{\hat{\sigma}, \hat{\alpha}, \hat{\gamma}, \hat{\rho}, \hat{\beta}, \hat{\theta}\}$ and $\hat{\boldsymbol{\delta}}$
2: Given $\{\hat{\boldsymbol{\delta}}, \hat{\alpha}, \hat{\beta}\}$, calculate new eq demand $\boldsymbol{u}'$ and transition of passengers $\tilde{\boldsymbol{A}}_{\boldsymbol{f}}$
3: Run the iteration process in Algorithm 1 to solve for new eq supplies $\boldsymbol{v}'$

Simulation Algorithm With Network Effects

Fix parameter values as estimates $\{\hat{\sigma}, \hat{\alpha}, \hat{\gamma}, \hat{\rho}, \hat{\beta}, \hat{\theta}\}$
Set initial guess of $\boldsymbol{\delta^0} = \hat{\boldsymbol{\delta}}$
Iterate from $k = 0$
    Given $\{\boldsymbol{\delta^k}, \hat{\alpha}, \hat{\beta}\}$, calculate new eq demand $\boldsymbol{u^k}$ and transition of passengers $\tilde{\boldsymbol{A}}_{\boldsymbol{f}}^{\boldsymbol{k}}$
    Run the iteration process in Algorithm 1 to solve for $\boldsymbol{v_f^k}$
    Plug $\{\boldsymbol{v_f^k}, \boldsymbol{u_f^k}, \hat{\boldsymbol{\xi}}, \hat{\theta}\}$ in to mean utility 6.2 and update $\boldsymbol{\delta^{k+1}}$
Stop until $\|\boldsymbol{\delta^{k+1}} - \boldsymbol{\delta^k}\| < \epsilon$
New equilibrium $\boldsymbol{v_f^*}, \boldsymbol{u_f^*}$

9.1 Eliminating surge multiplier

In this section, I want to study whether Uber's surge pricing improves matching efficiency across locations. Unlike fixed fare of taxis, Uber use surge multiplier to efficiently adjust drivers' search incentives of different locations. When some locations have higher demand than supply, Uber tends to charge a higher price than regular one to motivate more drivers to come. This higher price is product of surge multiplier and regular price. To investigate the effect of flexible pricing of Uber on matching efficiency, I eliminate Uber's surge multiplier such that all Uber's trips are calculated using the normal pricing structure. By comparing the new equilibrium with the benchmark, it could tell, to some extent, the role of flexible pricing in matching of this industry. The results are listed in table 8.

First of all, we can compare the new equilibrium without network effects to the baseline. After decline of Uber's prices, the aggregate searches of drivers do not change much for both taxi and Uber. Demand are more sensitive to this change such that Uber's demand increases by 8,150 (16%). Taxis' total demand decreases by 5,510 (1.9%) due to the price competition. As a result of demand change, taxis' pickups also declines slightly by 1.39% whereas Uber's pickups increases by 9.11%. We are more interested in comparing the frictions in the second panel. The type 1 friction (within market mismatches) of taxi decreases due to the decreased demand and supply. Uber's type 1 friction is not quite meaningful to discuss since it results from the error of using hyperbolic function to approximate perfect matching. Most interesting findings are the cross location mismatches. Taxi's cross location friction decreases by 1,894 (12.8%) whereas Uber's friction increases by 2,811 (152%) trips. The increased type 2 friction of Uber is worth \$ 30,455 fares. The decreased cross location mismatches

of taxis may result from the competition effect, since Uber's product becomes more competitive with lower price. Without the help of surge pricing, Uber's misallocation of drivers makes its matching less efficient. As for the last panel, taxis' profits decrease due to price competition by $ 30,300. Though Uber's demand increases due to lower price, its total revenue decreases by $ 57,990. The welfare gain of passengers measured by inclusive value of expected utility prior to choices and matching process is 511,140 which is worth $ 120,400.

Then, the last column reports equilibrium with network effects such that demand and supply levels will change not only prices but also mean utilities in passengers' choice problem. I find that both supply and demand of taxi and Uber change further more than without network effect in the same direction. For example, the total supply of taxis decreases by 3,240 which is more than previous case. The demand of Uber increases to 62,065 compared to 58,374. This finding reflects the positive feedback loop in two sided market. After Uber's price decreases, demand for Uber increases which further increases utility of choosing Uber through direct network effect and so forth. As a result, the total pickups of taxis decreases more and Uber's pickups increase more than without network effect. To conclude, existence of network effect could expand the effect of price drop on market share in this counterfactual.

As for frictions, taxis' within location friction decreases. This is mainly due to the further decreased demand and supply of taxis. Cross location mismatches of taxis and Uber have opposite results as well. Taxi's type 2 friction decreases by 2,573 and Uber's increases by 3,152. Finally, in the last panel, taxis' profits decreases slightly more than the case without network effect. Uber drivers make more money after allowing network effect because it make more demand for Uber to compensate the price drop. However, consumer welfare gain decreases from $ 120,400 to $ 96,977 due to network effects.

Table 8: Eliminating surge multiplier

supply, demand, match	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v^i_{yt}$	418,100	417,230	414,860
$\Sigma_{i,t} v^i_{xt}$	82,707	83,375	84,219
$\Sigma_{i,t} u^i_{yt}$	283,510	278,000	273,730
$\Sigma_{i,t} u^i_{xt}$	50,224	58,374	62,065
$\Sigma_{i,t} m^i_{yt}$	173,230	170,810	168,470
$\Sigma_{i,t} m^i_{xt}$	47,738	52,088	54,371
two type friction			
$within friction_y$	95,547	94,343	93,092
	\$ 1,286,400	\$ 1,272,300	\$ 1,256,800
$within friction_x$	635	758	786
	\$ 10,517	\$ 10,516	\$ 10,804
$cross friction_y$	14,738	12,844	12,165
	\$ 203,530	\$ 178,120	\$ 169,460
$cross friction_x$	1,850	4,661	5,002
	\$ 34,450	\$ 64,905	\$ 66,603
welfare			
$\$taxiprofit$	\$2,510,400	\$2,480,100	\$2,452,900
$\$Uberprofit$	\$779,350	\$721,360	\$744,330
$consumer\ welfare$	505,210	511,140	510,020
$\Delta\$consumer\ welfare$	NA	\$120,400	\$ 96,977
$\Delta\$social\ welfare$	NA	\$32,110	\$4,457

Note: $within friction_y$ is $\Sigma_{i,t} \min\{u^i_{yt}, v^i_{yt}\} - \bar{m}^i_{yt}$.
$cross friction_y$ is $\Sigma_t \min\{\Sigma_i \max\{u^i_{yt} - v^i_{yt}, 0\}, \Sigma_i \max\{v^i_{yt} - u^i_{yt}, 0\}\}$.

9.2 Improving traffic conditions

In this section, I simulate equilibrium after improving traffic conditions. To do so, I replace the trip time $\{\chi_t^{ij}\}_{\forall i,j,t}$ from location i to j in period t computed using 2016 sample by the trip time of the same route in 2010 (the year before Uber's entry)[36]. I have shown in figure 2 that traffic in 2010 is relatively faster than 2016. Thus, by solving new equilibrium with traffic condition in 2010 can suggest the role of traffic condition on matching efficiency. But, in this paper, I do not build any relationship between traffic condition and supply of vehicles. Thus, all counterfactuals only provide a partial effect or first order response of factor change on equilibrium. The new equilibrium and welfare in comparison to benchmark is provided in table 9.

First we can compare the efficiency change without network effects. Aggregate demand for taxi and Uber do not change in the new equilibrium[37]. Daily aggregate supply of both taxi and Uber increase. For example, total supply/searches of taxi drivers increase by 68,610 (16.4%) and Uber's increase by 10,963 (13.25%). As consequence, the total pickups of taxis increase by 7,110 (4.1%) and of Uber increase by 971 (2.03%). The two types of friction also changes. Both frictions of taxi decrease especially for the cross location mismatches. The total number of friction 2 of taxis decreases by 6,142 (41.67%) which is equivalent to \$ 80,620 trip fares. As for Uber, the within location friction is negligible due to perfect matching assumption. Uber's cross location mismatches also decrease by 1,014 trips and the loss of fare decreases by \$ 18,618. The total revenue of taxis increases by \$ 107,400 (4.2%) and of Uber increases by \$ 17,830 (2.28%). The welfare gain to consumers as measured by compensating variation (CV) is zero because inclusive value of their expected utility prior to matching does not change without network effect. However, consumer welfare after matching process changes.

Then, the last column of table 9 shows the equilibrium with network effect which means demand will respond to the change of supply. The total supply of taxis increases further than without network effect. Uber's supply also increases compared to the benchmark however less than without network effect. There could be two possible explanations for the smaller supply of Uber than without network effect. One reason is that choice decisions of passengers change after responding to network effect. As

[36]The trip time for any route ij at t of a representative weekday of 2010 can be calculated in the same way as 2016

[37]The marginal change is due to computation error between $\boldsymbol{\delta_f} \leftrightarrow \boldsymbol{u_f}$

consequence, not only the market share/demand changes, but also the destinations of Uber's passengers change such that Uber's passengers tend to travel a long distance. The other reason is that the search values of Uber changes due to the demand change and Uber drivers are more likely to search a location far away. Both reasons make Uber drivers spend more time on travelling than searching. Taxi's demand increases more than without network effect and Uber's demand drops. One reason is that taxis have stronger network effect than Uber and increased mean utility of choosing taxis is higher than Uber. As results, taxi's pickups increase by 18,470 (10.6%) which is quite larger than 4.1% without network effect. It is interesting that Uber's pickups decreases in new equilibrium rather than increase in the previous case which implies that we could even have opposite conclusions with or without network effect.

As for frictions, the type 1 friction of taxis increases both in numer and money value. It is because of both increased demand and supply of taxis, and the random matching assumption. The type 2 friction of taxis also decreases compared to benchmark but is slight larger than the case without network effect. The same finding applies to Uber's type 2 friction. This finding implies that network effects actually make the matching across locations less efficient in this counterfactual. I have discussed the ambiguous impact of network effects on matching efficiency across locations earlier in section 3.

At last, the total profits of taxi and Uber increase after traffic improvement. However, taxis' profits are higher than without network effect whereas Uber's profits are less than without network effect. Consumer welfare increases by $ 748,900 benefiting from traffic improvement via the network effects in utilities.

Table 9: Traffic Improvement

supply, demand, match	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v^i_{yt}$	418,100	486,710	498,920
$\Sigma_{i,t} v^i_{xt}$	82,707	93,670	91,857
$\Sigma_{i,t} u^i_{yt}$	283,510	283,030	300,120
$\Sigma_{i,t} u^i_{xt}$	50,224	50,182	48,391
$\Sigma_{i,t} m^i_{yt}$	173,230	181,340	191,700
$\Sigma_{i,t} m^i_{xt}$	47,738	48,709	46,681
two type friction			
$within friction_y$	95,547	93,090	98,923
	$ 1,286,400	$ 1,252,900	$ 1,328,200
$within friction_x$	635	635	611
	$ 10,517	$ 10,531	$ 10,406
$cross friction_y$	14,738	8,596	9,501
	$ 203,530	$ 122,910	$ 134,920
$cross friction_x$	1,850	836	1,097
	$ 34,450	$ 15,832	$ 21,807
welfare			
$\$taxiprofit$	$2,510,400	$2,617,800	$2,741,500
$\$Uberprofit$	$779,350	$797,180	$784,460
$consumer\ welfare$	505,210	505,210	536,070
$\Delta\$consumer\ welfare$	NA	$0	$ 748,900
$\Delta\$social\ welfare$	NA	$ 125,230	$ 985,110

9.3 Regulating Uber's supply

In this counterfactual, I study the proposed regulatory policy of NYC government on Uber in summer 2015. At that time, the number of Uber's affiliated vehicles is growing at a monthly rate of 3%. The total number of licensed Uber vehicles was 26,000 which outnumber the total medallions, though not all Uber vehicles are on the street together. Since the government blames Uber for contributing to traffic congestion, it proposed to restrict the growth of Uber to 1% annually. To simulate the result of this policy, I decrease the assumed total number of Uber vehicles in my model by 30%. This 30% comes from a simple calculation of ten months growth rate $3\% * 10$ from mid 2015 to my sample April 2016. Similarly, I simulate two equilibria with or without network effect in this scenario. The results are in table 10.

First, compare the second and the first column. After dropping 30% Uber vehicles, the total supply of Uber cars decrease by 17,377 (21%). The total number of taxis' supply does not change without network effect. The reason is that demand for taxis in this case does not change due to fixed mean utility. Given the unchanged distribution of demand for taxis, the equilibrium supply of taxis does not change as well. Similarly, demand for Uber is also unchanged. But due to decline of Uber's supply, total pickups of Uber decrease by 3,387 (7.1%). As for the frictions, we only need to compare Uber's type 2 friction. Uber's cross location mismatches increases by 3,020 which is worth $ 56,702 fares. The increase is because of unchanged Uber demand and the decreased supply of Uber. However, demand of passengers should respond to change of supply as it affects matching probability and waiting time. The importance of accounting for network effect is reflected in this example. Finally, taxis are not affected in this case without network effect. The total profit of Uber decrease by $ 63,250.

Next, compare the last column that allows network effect with first two columns. Total supply of taxis does not change much but its demand increases by 3,790. However, the decline of Uber's demand is 6,353 (12.6%) which is greater than increased taxi's demand. The total pickups of taxi increase by 1,660 and pickups of Uber decrease by 7,004 more than without network effect. The difference is because that utility of choosing Uber declines due to less supply of Uber cars. As for frictions, taxis type 1 friction increases a bit due to increased demand for taxis. Moreover, its type 2 friction also increases from 14,738 to 15,595. Though the difference is small, it reflects the direction that taxis' matching friction may go if there is less competition from Uber. Uber's cross location mismatches is small than without network effect as expected,

but it is still greater than benchmark. In terms of profits, taxis make $ 34,300 more money in a day shift after regulating Uber compared to $ 119,320 profit loss of Uber. Moreover, passengers are worse off by $ 98,829 after this regulation.

Table 10: Cap Uber's growth

supply, demand, match	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v^i_{yt}$	418,100	418,120	417,610
$\Sigma_{i,t} v^i_{xt}$	82,707	65,330	64,622
$\Sigma_{i,t} u^i_{yt}$	283,510	283,030	287,300
$\Sigma_{i,t} u^i_{xt}$	50,224	50,182	43,871
$\Sigma_{i,t} m^i_{yt}$	173,230	173,070	174,890
$\Sigma_{i,t} m^i_{xt}$	47,738	44,351	40,734
two type friction			
$within friction_y$	95,547	95,596	96,812
	$ 1,286,400	$ 1,287,100	$ 1,304,800
$within friction_x$	635	700	602
	$ 10,517	$ 11,670	$ 9,994
$cross friction_y$	14,738	14,367	15,595
	$ 203,530	$ 198,620	$ 215,750
$cross friction_x$	1,850	4,870	2,534
	$ 34,450	$ 91,152	$ 46,764
welfare			
$\$taxiprofit$	$2,510,400	$2,507,800	$2,544,700
$\$Uberprofit$	$779,350	$716,100	$660,030
$consumer\ welfare$	505,210	505,210	501,540
$\Delta\$consumer\ welfare$	NA	$0	$ -98,829
$\Delta\$social\ welfare$	NA	$-65,850	$-183,849

10 Conclusion

In this paper, I study the role of network effects on matching friction with an application to taxi and Uber drivers searching for passengers in New York City. To be specific, I model Uber and taxi drivers' dynamic search decisions among 40 defined locations in NYC and passengers' static choice decision. I focus on a day shift of a representative weekday in April 2016. In this industry, due to the fixed pricing structure of taxis, market is not cleared in prices leaving spatial mismatches across locations such that some areas have waiting passengers (excess demand) and some have vacant cars (excess supply). In additional of inefficient pricing, I also add network effects in the model and study its impact on matching efficiency. If we treat a market (location-time) as a platform, network effects exist if there are externalities between drivers and passengers when they make search decisions in this market (platform). The source of network effects could be passengers' preference to high matching probability and short waiting time, and drivers' preference to hight matching probability and less vacant time. I allow both direct and indirect network effects in demand side by adding supply and demand into utility function of passengers.

Estimates of the model show the existence of network effects and positive feedback loop between drivers and passengers. There is also positive externalities among passengers. Then, I study how network effects affect spacial mismatches of drivers and passengers in counterfactuals and whether ignoring it would cause inaccurate conclusions. I simulate three counterfactual scenarios with the model. First, I study whether Uber's surge pricing improves matching efficiency by eliminating it. I find a significant increase of Uber's mismatches without surge pricing. However, lower price without surge pricing would make more pickups for Uber by 6,633 in total of a daytime. Second, I study to what extent traffic condition matters for matching efficiency by using the traffic speed in 2010. In the new equilibrium of better traffic, both Uber and taxis have less mismatches by 33.71% and 40,7% respectively. Their profits also increase. However, without considering network effect, Uber's pickups are predicted to increase whereas it decreases with network effects. Third, I study the regulatory policy of capping Uber's number of vehicles. I reduces number of Uber cars by 30%. I find that taxis' pickups increases by 1,660 far less than the decrease of Uber's pickups 7,004. Interestingly, due to less competition, taxis' mismatches increases. Uber's matching efficiency decreases. All counterfactuals show different results either in magnitude or

sign for whether network effects are considered.

References

[1] Amstrong, M. (2006):"Competition in two-sided markets," *RAND Journal of Economics*, 37(3),668-691.

[2] Ackerberg and Gowrisankaran (2006)"Quantifying Equilibrium Network Externalities in the ACH Banking Industry,"*RAND journal of economics*, 37(3),738-761.

[3] Buchholz, N.(2016):"Spatial Equilibrium, Search Frictions and Efficient Regulation in the Taxi Industry," working paper.

[4] Butters, G. R. (1977):"Equilibrium distributions of sales and advertising prices"*Review of Economic Studies*, 465-491.

[5] Burdett, K., S. Shi, and R. Wright (2001):"Pricing and Matching with Frictions,"*Review of Economic Studies*, 465-491.

[6] Blanchard, O. and P. Dianmond (1989):"The Beveridge Curve,"*Brookings Papers on Economic Activity*, 1989, 1-76.

[7] Chu, J. and Manchanda, P.(2016):" Quantifying cross and direct network effects in online consumer-to-consumer platforms",*Marketing Science*, 35(6), 870-893.

[8] Chou, C. and Shy,O. (1990):"Network effects without network externalities," *International Journal of Industrial Organization*,8, 259-270.

[9] Church, J, and Gandal, N. (1992):"Network effects, software provision, and standardization," *Journal of Industrial Economics*, 40, 85-103.

[10] Clements,M. T., and H. Ohashi(2005):"Indirect network effects and the product cycle: U.S. video games, 1994-2002." *Journal of Industrial Economics*, 53(4),515-542.

[11] Corts, K. S., and M. Lederman (2009):"Software exclusivity and the scope of indirect network effects in the US home video game market." *International Journal of Industrial Organization*, 27(2),121-136.

[12] Crawford, V.P. and J. Meng(2011):"New York City Cab Drivers' Labor Supply Revisisted: Reference-dependent Preferences with Rational Expectations Targets for Hours and Income"*American Economic Review*,101,1912-1932.

[13] Chen,K.,J. Chevalier, P.Rossi and E. Oehlsen(2017):"The Value of Flexible Work: Evidence from Uber Drivers,"working paper.

[14] Chen,K. and M. Sheldon(2015):"Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform" working paper.

[15] Dubé, J.-P. H., G. J. Hitsch, and P.K. Chintagunta (2010):" Tipping and Concentration in Markets with Indirect Network Effects",*Marketing Science*, 29(2),216-249.

[16] Frechette, G., A. Lizzeri, and T. Salz(2017):"Frictions in a Comeptitive, Regulated Market: Evidence from Taxis," Working paper.

[17] Farber,H.S.(2005):"Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers",*Journal of Political Economy*,113,46-82.

[18] Farber,H.S.(2008):"Reference Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers"*American Economic Review*, Juen 2008, 1069-1082.

[19] Farrell, J., and G. Saloner (1986):"Installed base and compatibility: innovation, product preannouncements, and predation," *American Economic Review*, 940-955.

[20] Goolsbee, A. and Klenow, P.(2002):" Diffusion of home computers",*Journal of Law and Economics*, 45, 317-343.

[21] Gandal, N., Kende, M. and Rob, R. (2000):"The dynamics of technological adoption in hardware/software systems: the case of compact disk players," *RAND Journal of Economics*, 31,43-62.

[22] Gandal, N. (1994):"Hedonic price indexes for spreadsheets and an empirical test of network externalities," *RAND Journal of Economics*, 25,161-170.

[23] Kreindler, G.(2018):"The welfare effect of road congestion pricing: experimental evidence and equilibrium implications" working paper.

[24] Katz, M.L.m and C.Shapiro (1985):"Network externalities, competition, and compatibility," *American Economic Review*, 424-440.

[25] Lagos, R.(2000):"An Alternative Approach to Search Frictions,"*The Journal of Political Economy*, 108,851-873.

[26] Luo, R.(2016):“ Network Effect and Multi-Network Sellers’ Dynamic Pricing: Evidence from the US Smartphone Market” working paper.

[27] Lee, R. S.(2013):“ Vertical Integration and Exclusivity in Platform and Two-sided Markets”,*American Economic Review*, 103(7),2960-3000.

[28] Mortensen, D.T. and C.A. Pissarides (1999):“New Developments in Models of Search in the Labor Market,” *Handbook of Labor Economics*, 3,2567-2627.

[29] Nair, H., P. Chintagunta, and J.-P. Dubé(2004):“Empirical analysis of indirect network effects in the market for personal digital assistants,” *Quantitative Marketing and Economics*, 2(1),23-58.

[30] Ohashi, H.(2003):“The Role of Network Effects in the US VCR Market, 1978-1986.”*Journal of Economics and Management Strategy*, 12(4),447-494.

[31] Park, S.(2004):“Quantitative Analysis of Network Externalities in Competing Technologies: The VCR case,” *Review of Economics and Statistics*, 86(4),937-945.

[32] Pissarides, C. A. (1984):“ Search Intensity, Job Advertising, and Efficiency,”*Journal of Labor Economics*, 128-243.

[33] Qi, S.(2013):“The Impact of Advertising Regulation on Industry: the Cigarette Advertising Ban of 1971,”*RAND Journal of Economics*, 44, 215-248.

[34] Rochet, J.-C. and Tirole, J. (2003):“Platform competition in two-sided markets,” *Journal of the European Economic Association*, 1(4).

[35] Rysman, M.(2004):“Competition between Networks: A Study of the Market for Yellow Pages.”*Review of Economic Studies*,71,483-512.

[36] Saloner, G. and Shepard, A. (1995):“Adoption of technologies with network effects: an empirical examination of the adoption of automated teller machines,” *RAND Journal of Economics*, 26,479-501.

[37] Saeedi, M.(2014):“Reputation and Adverse Selection: Theory and Evidence from eBay,”*working paper*.

[38] Shapiro, M. H.(2018):“Density of demand and the benefit of Uber,” Working paper.

[39] Weintraub, G., C.Benkard, P. Jeziorski, and B. Van Roy(2008):"Markov Perfect Industry Dynamics with Many Firms," *Econometrica*, 76,1375-1411.

[40] Weintraub, G., C.Benkard, P. Jeziorski, and B. Van Roy(2008):"Nonstationary Oblivious Equilibrium," Manuscript.

[41] Xu, Y. (2008):"A structural empirical model of R&D, firm heterogeneity, and industry evolution,"*Journal of Industrial Economics*, forthcoming.

Appendix

Proof of proposition 1 In order to have one island with excess supply and one island with excess demand in equilibrium, the model parameters should have two properties. First, the price in island 1 is greater than island 2, $p_1 > p_2$. This is also condition in Lagos (2000) such that one island is more profitable. Second, the population in island 2 must be greater to offset or even dominate the negative effect of low supply on demand. This second condition requires that islands are not only heterogeneous in price but also in market size such that d_2 of island 2 is large enough to make $u_2 > v_2$. Given $p_1 > p_2$, suppose in equilibrium $v_1^* > u_1^*$ and $v_2^* < u_2^*$. By condition E3, we have $m_1^* = u_1^*$ and $m_2^* = v_2^*$. Plugging equilibrium matches in to condition E1, we have $v_1^* = \frac{p_1}{p_2} u_1^*$. Then solving demand equation of island 1, we can get explicit form of equilibrium supply and demand in island 1:

$$
\begin{aligned}
v_1^* &= \frac{p_1}{p_2} u_1^* \\
u_1^* &= \frac{d_1}{1 + \alpha - \beta \frac{p_1}{p_2}}
\end{aligned}
$$

We can solve equilibrium supply in island 2 by condition E4. Then plugging supply in island 2 into demand equation in island 2 solves equilibrium demand in island 2 which are:

$$
\begin{aligned}
v_2^* &= N_y - v_1^* \\
u_2^* &= \frac{\beta}{1 + \alpha} v_2^* + \frac{d_2}{1 + \alpha}
\end{aligned}
$$

In order to sustain this equilibrium, parameter values of $\{p_i, d_i, N_y, \alpha, \beta\}$ need to satisfy the following conditions. First condition is by default such that demand is positive. The second condition needs to make island 1 more profitable than island 2. The third inequality guarantees demand is positive in equilibrium if thick market effect β is relatively smaller than congestion or direct network effects α. The last inequality derives from excess demand in island 2. It requires that d_2 need to be large enough

such that $u_2^* > v_2^*$ in equilibrium.

$$
\begin{aligned}
d_1, d_2 &> 0 \\
p_1 &> p_2 \\
1 + \alpha - \beta\frac{p_1}{p_2} &> 0 \\
u_2^* - v_2^* = \frac{d_2}{1+\alpha} + (\frac{\beta}{1+\alpha} - 1)(N_y - \frac{p_1}{p_2}u_1^*) &> 0
\end{aligned}
$$

Proof of proposition 2 The demand equation can be rewritten as in (3.4):

$$
\begin{aligned}
u_{yi} &= -\alpha u_{yi} + \beta v_{yi} + \underbrace{\gamma u_{xi} - \theta v_{xi} + d_{yi}}_{D_{yi}} \forall i \\
u_{xi} &= -\alpha u_{xi} + \beta v_{xi} + \underbrace{\gamma u_{yi} - \theta v_{yi} + d_{xi}}_{D_{xi}} \forall i
\end{aligned}
$$

We can deem D_{yi} as d_i in proposition 1 above. The equilibrium in which taxis have excess supply in island 1 and excess demand in island 2, while Uber has excess supply in both islands satisfies:

$$
v_{y1}^* = \frac{p_{y1}}{p_{y2}} u_{y1}^* \tag{T1}
$$

$$
u_{y1}^* = \frac{D_{y1}}{1 + \alpha - \beta\frac{p_{y1}}{p_{y2}}} \tag{T2}
$$

$$
v_{y2}^* = N_y - v_{y1}^* \tag{T3}
$$

$$
u_{y2}^* = \frac{\beta}{1+\alpha} v_{y2}^* + \frac{D_{y2}}{1+\alpha} \tag{T4}
$$

for taxis. These four equations are obtained similar to proposition 1. The only different is that D_{yi} contains supply and demand of opponent firm Uber within the same island i. As for Uber, assuming equal prices of Uber in both island, we have $\frac{u_{x1}^*}{v_{x1}^*} = \frac{u_{x2}^*}{v_{x2}^*}$ such that drivers' probabilities of matching are same in both islands. Together with Uber's demand equations, we can solve expression for Uber's demands and supplies as below.

$$v_{x1}^* = \frac{D_{x1}}{D_{x1} + D_{x2}} N_x \tag{X1}$$

$$u_{x1}^* = \frac{\beta}{1+\alpha} v_{x1}^* + \frac{D_{x1}}{1+\alpha} \tag{X2}$$

$$v_{x2}^* = \frac{D_{x2}}{D_{x1} + D_{x2}} N_x \tag{X3}$$

$$u_{x2}^* = \frac{\beta}{1+\alpha} v_{x2}^* + \frac{D_{x2}}{1+\alpha} \tag{X4}$$

Given the eight equations (T1-T4, X1-X4) and defined D_{fi}, we can solve eight unknowns $\{u_{fi}^*, v_{fi}^*\}$ after some algebra. To make such equilibrium exist, the solved equilibrium and demands as explicit form of parameters $\{N_f, d_{fi}, p_{yi}, \alpha, \beta, \gamma, \theta\}$ must satisfy $u_{y2}^* > v_{y2}^*$ and $u_{xi}^* < v_{xi}^*, \forall i$[38]. Without solving the explicit forms, the intuition for existence of such equilibrium is as follows. First, excess supply of Uber in both islands can be achieved when N_x is large enough. For example, $v_{x1}^* - u_{x1}^* = (1 - \frac{\beta}{1+\alpha})v_{x1}^* - \frac{D_{x1}}{1+\alpha}$ with $D_{x1} \equiv \gamma u_{y1} - \theta v_{y1} + d_{x1}$. Larger N_x makes both v_{x1}^* and u_{x1}^* larger. As long as u_{y1}^* and D_{x1} do not increase as fast as increase of v_{x1}^*(i.e. $\gamma = 0$), difference $v_{x1}^* - u_{x1}^*$ increase in N_x. Second, to guarantee taxis' demand in island 2 large than supply $u_{y2}^* - v_{y2}^* > 0$, it still requires d_{y2} large enough as in proposition 1. Large d_{y2} increases demand of taxis in island 2 as show in T4 such that demand exceeds supply. Moreover, increased demand u_{y2} will not affect Uber's equilibrium conditions X1-X4 through D_{x2} as long as γ is small or even equals zero.

[38] Excess supply of taxi in island 1 is guaranteed given $p_{y1} > p_{y2}$ and condition T1.

Congestion Pricing

In this counterfactual, I study the outcomes of congestion pricing that NYC government proposes in recent days. This policy intends to reduce the traffic congestion in CBD Manhattan by charging surcharge to both taxis and Uber. To be specific, for any trip touching Manhattan south of 96th Street, taxis riders need to pay $2.5 surcharge and Uber riders need to pay $2.75 surcharge. I add these surcharges to the prices and simulate the results in table 11.

The first panel shows that demand for both taxis and Uber decreases due to the increased trip prices. So do the total pickups of Uber and taxis. In the second panel, I find that the cross location mismatches decrease after congestion pricing. The reason is that, demand for trips in Manhattan CBD decreases in prices. As a result, the expected profit of drivers decreases in Manhattan which in return makes oversupply in Manhattan decline. The changes of aggregate supply/demand in CBD before and after the policy are displayed in figure 16. Supply of both firms decline slightly with CBD surcharge. Finally, drivers' profits and passengers are all worse off if government implements congestion pricing in CBD of Manhattan.

Table 11: Congestion pricing

supply, demand, match	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v^i_{yt}$	418,100	413,760	412,190
$\Sigma_{i,t} v^i_{xt}$	82,707	81,083	80,736
$\Sigma_{i,t} u^i_{yt}$	283,510	269,440	266,840
$\Sigma_{i,t} u^i_{xt}$	50,224	45,678	44,768
$\Sigma_{i,t} m^i_{yt}$	173,230	166,100	164,490
$\Sigma_{i,t} m^i_{xt}$	47,738	43,831	42,922
two type friction			
$within friction_y$	95,547	91,069	90,210
	$ 1,286,400	$ 1,399,600	$ 1,386,900
$within friction_x$	635	570	557
	$ 10,517	$ 10,921	$ 10,726
$cross friction_y$	14,738	12,272	12,142
	$ 203,530	$ 189,160	$ 187,200
$cross friction_x$	1,850	1,273	1,288
	$ 34,450	$ 26,714	$ 27,346
welfare			
$\$taxiprofit$	$2,510,400	$2,435,600	$2,416,000
$\$Uberprofit$	$779,350	$737,510	$727,070
$consumer\ welfare$	505,210	467,660	461,270
$\Delta\$consumer\ welfare$	NA	$-715,550	$ -857,030
$\Delta\$social\ welfare$	NA	$ -832,190	$ -554,990

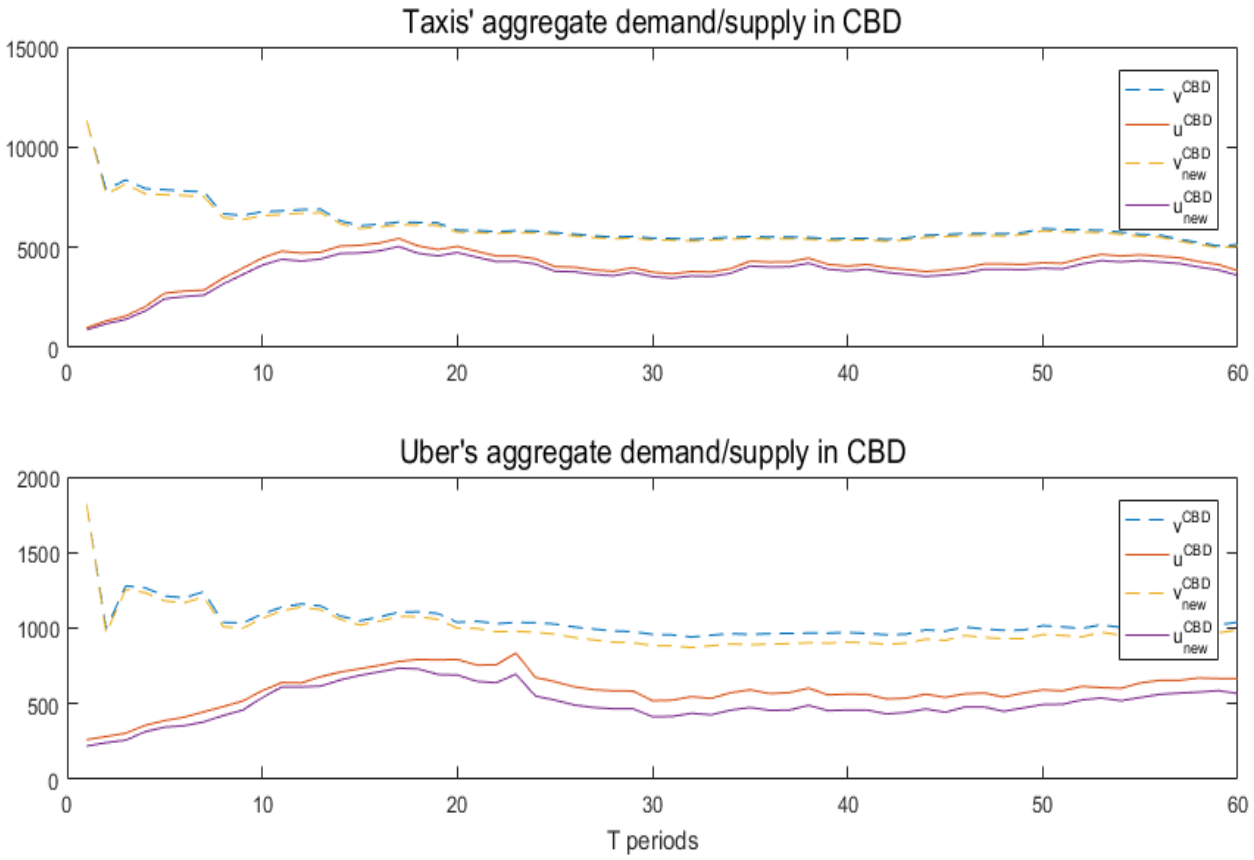
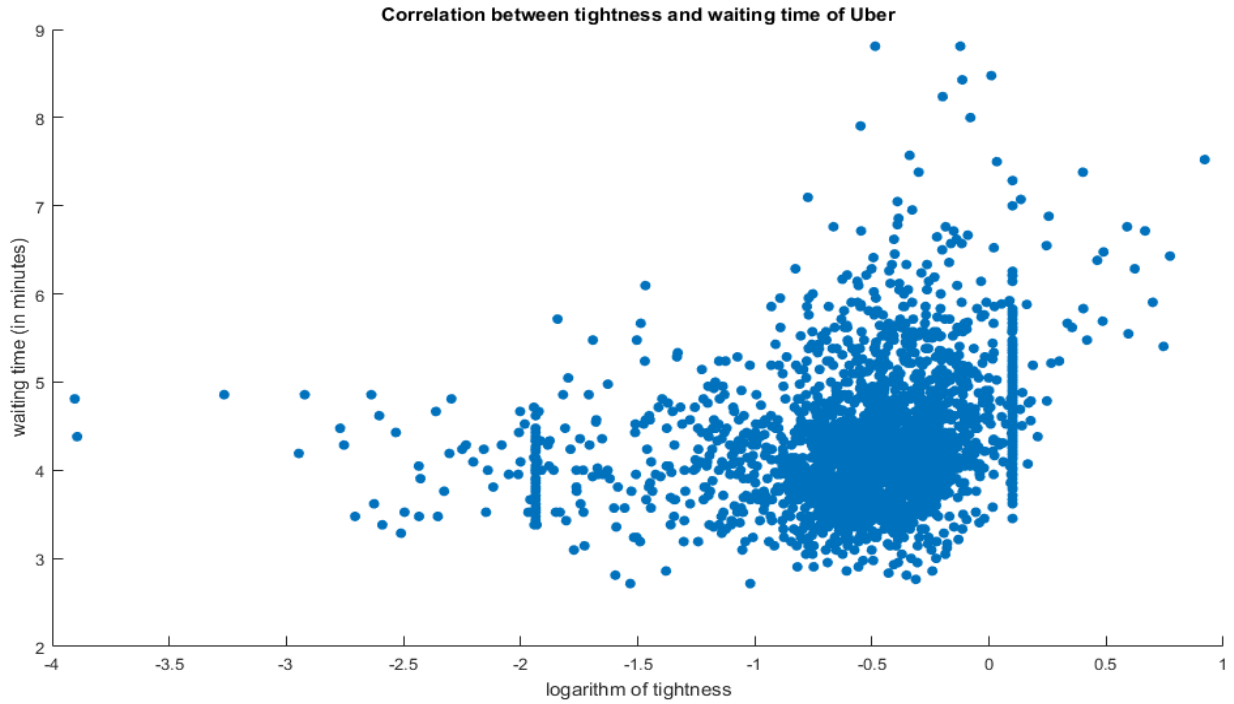
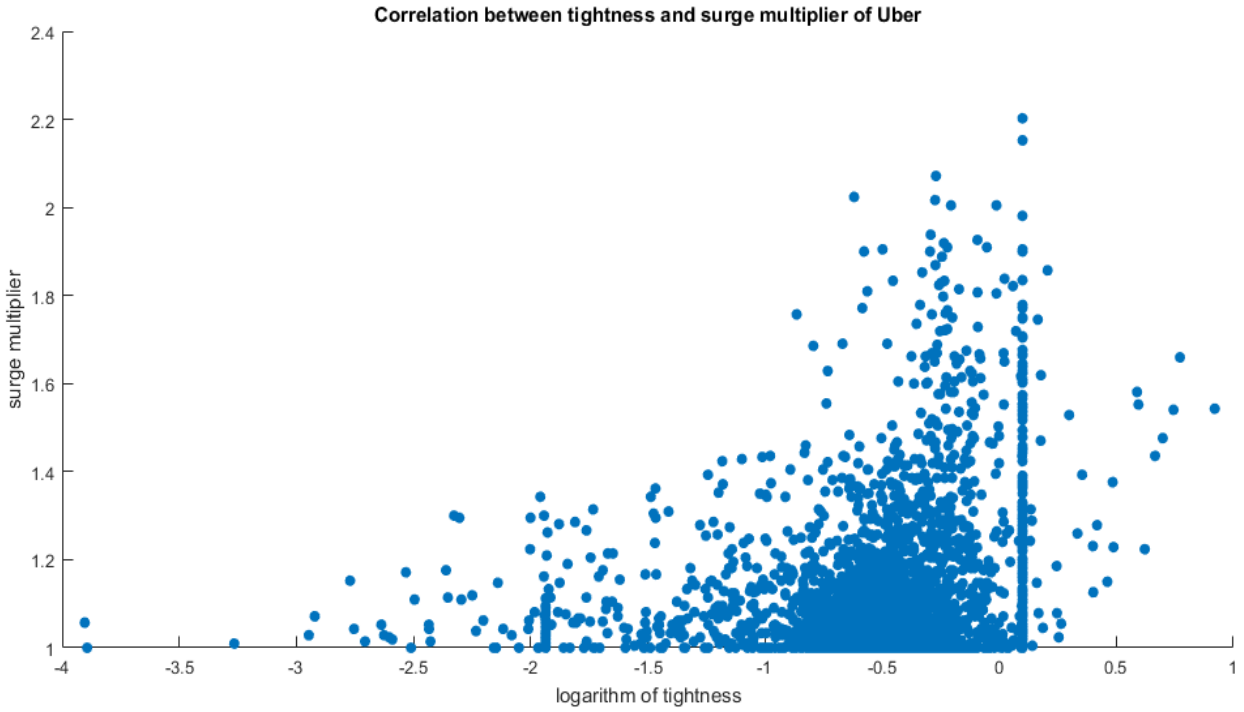


Figure 16: Aggregate demand/supply in CBD over time


Correlation between tightness and waiting time of Uber
waiting time (in minutes)
logarithm of tightness
9
8
7
6
5
4
3
2
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1


Correlation between tightness and surge multiplier of Uber
surge multiplier
logarithm of tightness
2.4
2.2
2
1.8
1.6
1.4
1.2
1
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1