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

Bo Bian

Pennsylvania State University

January 2, 2018

Abstract

In this paper, I model the search and matching process among drivers and passengers of taxis and Uber cars 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 and passengers make static discrete choice decisions among taxi and Uber. I model network effects by adding demand and supply levels to both sides' decisions in the form of a matching probability or waiting time. I use the nonstationary oblivious equilibrium concept to solve the model and analyze equilibrium frictions as mismatches between empty cars and waiting passengers. I show that network effects and supply competition, in addition to fixed fare and search costs, have extensive effects on frictions and welfare in three counterfactual scenarios. The first decreases Uber supply by 30 percent, the second improves traffic conditions and the third eliminates the Uber surge multiplier. I find that taxis' pickups increase by 5.9 percent if traffic improves but do not increase significantly under supply regulation. Taxis' profits increase by 1 percent under supply regulation and increase by 5.18 percent if traffic improves. Uber's search friction increases after eliminating the surge multiplier or restricting supply. Consumer welfare decreases if Uber supply is restricted. Without network effects, search frictions and pickups will be underestimated.

Key Words: spatial competition, network effects, search and matching, regulation.

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. The matching function returns the number of successful trades given the numbers of searchers on both sides of the market. However, the source of frictions underling such a function is not explicitly modelled.

In more recent literature, the microfoundations of the matching function are studied. Lagos (2000) builds a model of taxis' spatial search for passengers and finds that even without imperfection information and random search assumptions, aggregate mismatches over locations arise endogenously as outcomes of drivers' optimal search decisions. Specifically, when one location is better 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 that forms heterogeneity in profitability across locations. However, both Lagos (2000) and Buchholz (2016) focus on the prices as a source of spacial mismatches and study the supply side of the matching process.

Taxi industry is well known for existence of matching frictions such that some areas have excess demand whereas some have excess supply. I develop a model to analyze search and matching process of taxi industry following Buchholz (2016). This industry is ideal for analyzing search and matching frictions for several reasons. First, search decisions are made by individuals from both sides of the market such that there is no coordinator to help the market clear. Second, the taxi market is highly regulated, for example fixed fares and medallion numbers, which could cause frictions in market clearing. Third, there is no preference heterogeneity such that drivers and passengers are identical. For example, drivers cannot choose passengers or refuse a ride.

In this paper, I contribute to previous literature by developing a richer demand side with network effects in the sense that decisions of both suppliers and demanders depend on the size of the other side of the market. I study how this non-price factor influences matching efficiency of the market. On the supply side, drivers make spatial search decisions among locations by choosing the one that gives him 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 negatively affects expected profit through competition. 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 affects the matching probabilities or waiting time which affects their utility and choice decisions[1]. This interdependence between demand and supply has effects on matching efficiencies. For instance, if the sensitivity of demand to supply is positive, an increase of supply in a location with excess supply will increase demand which will further attract 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 taxis' search decisions alone as Buchholz (2016) does, I further allow the model to include a competitor Uber. Taxis or Uber cars are different products to passengers. Taxis and Uber compete for passengers in each location by providing products with different prices and qualities (supply). I also allow the matching technology to differ between taxis and Uber. The benefits of modelling Uber are twofold. First, in difference to congestion effect that decreases matching probability within taxis, competition from Uber affects taxis' profits and search decisions by changing demand for taxis. Second, there are real world debates about the influence of Uber on traditional taxis' profits and the regulation of Uber. My duopoly model can provide predications on market outcomes of regulatory policies or firm strategy.

To empirically analyze the model, I use data on trip records of taxis and Uber in April 2016 from the New York City Taxi and Limousine Commission(TLC). This dataset provides detailed information on taxis and Uber cars including pickup location and time for each trip. However, one important data limitation is that we do not observe the underlying supply and demand prior to matching process in each market. The pickups we observe in the data are successful matches. For Uber, I assume perfect matching such that pickups reflect demand or supply of the market. For taxis, I

[1] I refer to the positive externality between demand and supply as (indirect) network effects and the negative externality from the same side as congestion for the rest of this paper.

use Buchholz's strategy to find the equilibrium supply and demand imposed by our dynamic spatial search model which generates the same distribution of pickups in our model to data. 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.

During my sample month, there are 13 million pickups for taxis and 4.6 million for Uber cars. The total number of unique Uber cars that complete these trips is approximately 26 thousand which outnumbers the 13 thousand taxi medallions. Since Uber's entry to New York City in 2011, it has grown rapidly in number of affiliated cars and pickups. For example, in April 2015, Uber had only 12 thousands cars and 1.8 million monthly pickups. Due to the competition from Uber, the value of a taxi medallion has dropped. For example, the auction price of an independent unrestricted medallion dropped from \$ 0.7 million in 2011 to \$0.5 million in 2016. In order to protect the rent of taxis, in summer 2015, the NYC government proposed to cap the growth of Uber from 3% at monthly rate to 1% annually. This policy may affect passengers' welfare via waiting time and likelihood to get a ride which can be studied by my model.

As implication of my model, I simulate three counterfactuals and compare welfares with and without network effects. In the first counterfactual, I decreases Uber supply by 30%[2]. The second improves traffic conditions. In the third, I eliminate Uber's surge multiplier. I find that after restricting Uber's supply: (1) the demand and pickups of taxis do not increase significantly; (2) taxis' profits increase by 1%; (3) Uber's pickups and profits decrease significantly; (4) consumers' welfare decrease. In the case of traffic improvement: (1) both Uber and taxis' searches/supply increase; (2) taxis' profits increase by 5.18% and Uber's by 7%; (3) consumer welfare increases by 2.3%. After eliminating the Uber surge multiplier, cross locations mismatches of Uber increase but demand and consumer welfare increase due to price drop. Measurement of frictions, profits and consumer welfare differ with or without network effects.

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 effect matters for matching efficiency. Section 4 and 5 introduce industry background and

[2]Assuming that Uber grows at a fixed monthly rate of 3%, if the policy of capping Uber was implemented in June 2015, there were roughly 30% less Uber cars in April 2016.

the data I use for empirical model. Section 6 presents the empirical model. Estimation strategy and results are shown in sections 7&8. In section 9, I analyze three counterfactuals with my estimates. Section 10 concludes this paper with some discussion.

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 direct and indirect network effects. The direct network effects is that passengers' utility (or drivers' profits) is directly affected by the number of consumers (or drivers) from the same side of the market. For example, more passengers willing to take rides will reduce all passengers' probability to match a car. The indirect network effects refer to positive feedback loop between passengers and drivers. More drivers will increase passengers' matching probability and reduce search and waiting costs, and more demand results in more driver visits. Most recent empirical work on network effects focuses on indirect network effect of two-sided platform rather than direct network effects. Moreover, there is no industry studied in previous literature that composes of both direct and indirect network effects as in this paper. Without particular notice, I refer to indirect network effect as network effect and refer to direct network effect as congestion. There are many literature on network effects. Rysman (2004) estimates the network effects in Yellow Pages market and how it is related to market concentration. Dubé, Hitsch, and Chintagunta (2010) study network effects in video game market and its tipping effects. Similarly, Luo (2016) studies network effects in smartphone industry and how it affects carriers' dynamic pricing strategy. Lee (2013) studies software exclusivity in game industry with network effects. Other empirical network effects literature include: Ohashi (2003) and Park (2004) in the VCR market; Gowrisankaran, Park and Rysman (2014) in DVD player industry; Ackerberg and Gowrisankaran (2006) in ACH banking industry.

This paper also contributes to the search and matching literature by adding network effects. 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 such 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 one location is more lucrative than other locations, drivers will overcrowd that location leaving other locations with unserved passengers. Coexistence of excess demand and excess supply reflect frictions in the aggregate matching function. Buchholz (2016) extends Lagos's model 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 two-part tariff taxi fares prevent 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 to incorporate network effects. Second, I model cross-firm competition between taxi and Uber drivers via passengers' discrete choice demand. 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.

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 both 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 some other work related to this paper. Frechette, Lizzeri and Salz (2016) studies the effects of entry restrictions and matching frictions in NYC taxi industry. Chen et al.(2017) studies flexible labor supply of Uber drivers. Chen and Sheldon(2015) studies surge pricing and flexible work of Uber. Earlier work studying NYC taxi industry include Farber (2005, 2008), Crawford and Meng (2011).

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 is only taxis in the first model and 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 estimated in sections 6 to 8 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. 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[3]

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

The negative coefficient α on demand u_i measures congestion or negative externality of other passengers in the same island. The positive coefficient β on supply measures positive externality of supply on demand, the network effects. 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:

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

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

where m_i is matches in island i obtained from $m_i = \min\{u_i, v_i\}$ which implies perfect matching in each 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).

We 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:

$$v_1^* = \frac{p_1}{p_2} u_1^* \tag{supply in island 1}$$

$$u_1^* = \frac{d_1}{1 + \alpha - \beta \dfrac{p_1}{p_2}} \tag{demand in island 1}$$

$$v_2^* = N_y - v_1^* \tag{supply in island 2}$$

$$u_2^* = \frac{\beta}{1+\alpha} v_2^* + \frac{d_2}{1+\alpha} \tag{demand in island 1}$$

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 taxi pickups in equilibrium, $u_1^* + v_2^*$, from the minimum of total demand and supply, $\min\{u_1^* + u_2^*, v_1^* + v_2^*\}$. Our interest in this exercise is to do static comparative analysis of β on the friction $\min\{v_1^* - u_1^*, u_2^* - v_2^*\}$.

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 island 2 already had 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 $\frac{v_2^*}{1+\alpha} + \frac{\beta}{1+\alpha}\frac{dv_2^*}{d\beta}$. However, the first order derivative of excess demand w.r.t. β is positive because $\frac{v_2^*}{1+\alpha} + (\frac{\beta}{1+\alpha} - 1)\frac{dv_2^*}{d\beta} > 0$ and $\beta < 1 + \alpha$[4]. 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.

[4] $\beta < 1 + \alpha$ 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[5]. The parameters α and β measure congestion and network effect of each firm as in previous model. The new parameters γ and θ measure cross elasticity of demand to the supply and 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, we can 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, we 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

[5] 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 $\gamma = 0$[6]. 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 lost there. 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

[6] 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 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 and 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 when 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. One important notice is that FHV has open entry unlike taxis which have a fixed number of medallions. Black cars include ridesharing companies such as Uber, Lyft and Via[7]. 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 based 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[8]. 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 cover 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 Limousine Commission (TLC)[9]. The taxis trip records include all trips completed by

[7]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.

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

[9]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.

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 gap 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, we 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, we 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 location of the pick-up[10]. I identify the Uber trips according to its base numbers. 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. Out of all black car trips, Uber accounts for 72.6% and Lyft accounts for 11.6%. Thus, I only model Uber as competitor to taxis. 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 taxi 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[11].

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 supply 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 gives the surge multiplier of different types of Uber at that time and I use multiplier of UberX to calculate the data. Combining with trip records data by matching location and time, I can calculate approximate fare of Uber trips.

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

[11]Dropoffs data, if available, would help with model identification.

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[12]. The time period within a day in my sample is restricted to 6 a.m. to 4 p.m. In other words, 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 and exit[13] . 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.

[12] 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 day.

[13] 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

of weekdays, the number of active Uber drivers is fixed such that supply is tractable and most of them are full time Uber drivers. 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 each 10 minutes as a time period and 60 periods in total. Within each period, passengers and drivers randomly meet only once and successful contacts become pickups of that period. In other words, each driver only supplies once in a period[14]. I divide the city and select 40 geographic locations as my markets 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 markets 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. As discussed earlier, the pickups of a location-period are calculated as monthly average pickups over all weekdays of the same location-period.

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[15]. 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 minutes is transformed into number of 10-minute periods. For example, a 25 minutes trip takes 3 periods to complete. The trip distribution indicates the destination distribution of a given origin-period. This distribution can only be calculated for taxis since Uber's dropoffs are not in the data.

For any well defined market as a location-period combination, I construct market size and population trip distribution. Market size is widely used in discrete choice demand model to solve IIA assumption and to control substitution among inside products to outside option when price increases. In my demand model, the outside option

[14]Under this assumption, each driver only supplies once in a period which could make underestimated if the length of the defined period is long. For example, drivers could complete a trip within 10 minutes and pick up another passenger. Buchholz (2016) defines a five-minute period.

[15]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.

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[16].

Finally, I calculate the trip patterns of all travellers. I do not have data of all passengers' destination distribution. Instead, I calculate the trip patterns of taxis in November 2010 as proxy for population travelling patterns[17]. There are two implicit assumptions. First, I assume travellers paying full fare for subway follow the same travelling patterns to taxis' passengers in 2010. Second, the travelling patterns of 2010 and 2016 are the same after Uber's entry. The trip pattern is defined for any location in any period as shares of destinations. In other words, taxis' trip distribution of 2010 is deemed as travelling patterns of population in 2016 and trip distribution of taxis 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). We can observe the pickup distribution. In April 2016, taxis' monthly aggregate pickups during day shift 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 during day shift. Outer boroughs has 2.24% pickups in total. Comparing pickups of taxis between 2010 and 2016, we can observe huge decrease of pickups from 4.6 million to 3.7 million during day shift. 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

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

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

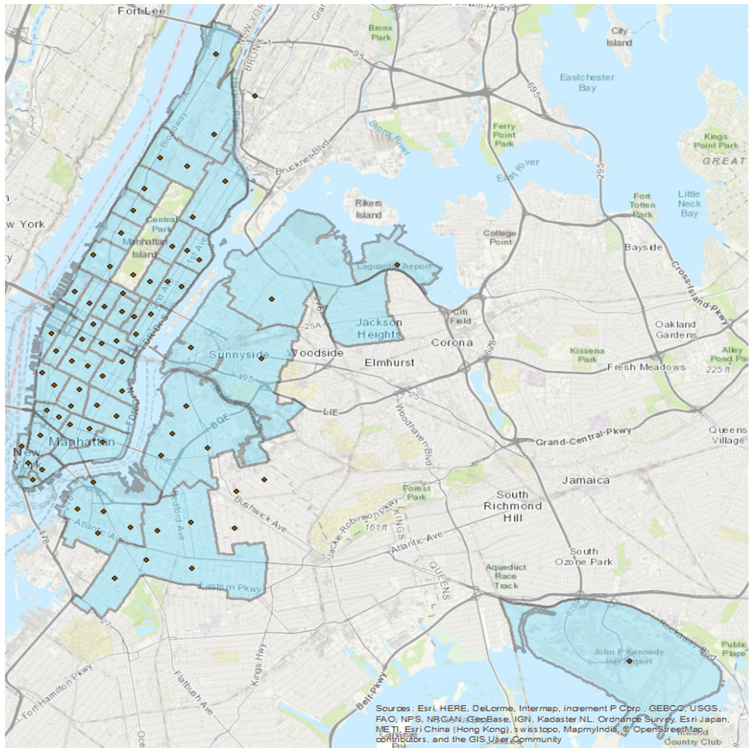


Figure 1: 40 select markets and 79 spots to collect Uber surge multiplier

changes after Uber's entry but not significantly. 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 destination distribution conditional on any given market. I use taxis' dropoffs data of November 2010 to approximate market demographics. We cannot use dropoffs in 2016 as market demographics because its endogenous outcomes of passengers' discrete choice among taxi, Uber and outside option. Table 2 provides a rough overview of what the dropoffs distribution looks like. 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 conditional on market can be discovered, which helps identify random coefficient of prices. For instance, given two markets with same prices, market size and supplies, differences in demands result from 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				

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 fare	96,000	17.68	7.92	28.4	10
Uber fare	96,000	22.39	9.12	37.95	12.48
taxi matches	2,400	72.28	8.80	163.02	57.34
Uber matches	2,400	20.37	8.33	34.95	10.25

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 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. In the next section, I will introduce the full structural model for estimation.

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 each side in a given market. When making supply/demand decision, the agent accounts for both the positive network effects from the other side of the market and negative congestion 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 do not fully contact with each other due to coordination failure. However, I assume perfect matching of Uber within the market. 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 frictional matching at the aggregate level.

Before delving into demand and supply decisions, I assume the timeline within a location-period is as follows:

timing of supply, demand and match of a market

Employed and unemployed Uber&taxi drivers arrive and become supply of the market.	Passengers make simultaneous discrete choice decision between Uber, taxi and outside option.	Matches are made between drivers of each firm and passengers within the market.	Unmatched passengers leave the market with subway.	Employed drivers deliver passengers and unmployed drivers make search decisions for next period.

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 previous period. Some of the cars either employed or unemployed are still on their way to the destination and will not necessarily arrive at a location in this period. All arriving cars become supply to that market in this period[18]. A passenger in this market has perfect information about fares and how likely he will find a taxi or Uber car[19]. 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 requested destinations and unemployed drivers choose locations to search next period.

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

[19] 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.

6.1 Passengers' Choice Problem

In a market defined by a location and time 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 willing to travel to j at period t choosing firm $f = x, y, o$ prior to matching process is:

$$U^{ij}_{cft,pre} = \tau^i_{ft} \exp\{U^{ij}_{cft,post}\} \tag{6.1}$$

where τ_{ft} is probability of being matched and $U^{ij}_{ft,post}$ is utility of ride conditional on being matched. The matching probability does not differ for different destinations j. The probability τ^i_{ft} depends on supply v^i_{ft}, demand u^i_{ft} and matches m^i_{ft} for Uber and taxis such that $\tau^i_{ft} = m^i_{ft}/u^i_{ft}$. Matches are a function of supply and demand $m^i_{ft} = m(v^i_{ft}, u^i_{ft})$ that will be discussed later. For outside option, the matching probability equals 1. Taking logarithm of (6.1) and specify posterior utility gives:

$$\begin{aligned}\log(U^{ij}_{cft,pre}) &= \log(\tau^i_{ft}) + U^{ij}_{cft,post} \\ &= \underbrace{\log(\tau^i_{ft}) + \theta \ln(u^i_{xt}/v^i_{xt})\mathbb{1}_{f=x} + d_x + d_i + t + \xi^i_{ft}}_{\delta^i_{ft}} + \alpha^{ij} \ln(p^{ij}_{ft}) + \varepsilon^{ij}_{cft}\end{aligned} \tag{6.2}$$

I specify utility conditional on taking a ride in the traditional way used in discrete choice models, as a linear equation of price p^{ij}_{ft}, product fixed effect d_x, market fixed effect d_i, time effect t, unobserved demand shock ξ^i_{ft} and idiosyncratic shock ε^{ij}_{cft}. I assume ε^{ij}_{ft} is an i.i.d. draw from a type I extreme value distribution. One special variable in (6.2) is the logarithm of the ratio of demand-to-supply. This is a measure of waiting time observed in Uber app or a measure of tightness. We can put all variables invariant to destination j together and denote as δ^i_{ft}. I allow price coefficients to differ across destinations j similar to random coefficient. I specify $\alpha^{ij} = \alpha_1 * \exp(\alpha_2 * distance^{ij})$ such that trips with different travel distances have different price elasticities.

The probability of a passenger conditional on travelling from i to j at time t choosing firm f is:

$$s_{ft}^{ij} = \frac{\exp(\delta_{ft}^{i} + \alpha^{ij}\ln(p_{ft}^{ij}))}{1 + \sum_{g=1,2}\exp(\delta_{gt}^{i} + \alpha^{ij}\ln(p_{gt}^{ij}))} \tag{6.3}$$

In the BLP model, we can estimate demand by matching choice probabilities of (6.3) to the market shares s_{ft}^{ij} obtained by dividing demand u_{ft}^{ij} by market size observed in the data. However, there are two obstacles to do this in this paper. First, I can only observe pickups m_{ft}^{i} rather than demand u_{ft}^{i}. Second, even though I can calculate demand for taxis u_{yt}^{ij} conditional on $\{i, j, t\}$ from destination distribution of taxis, Uber's destination distribution is not available in order to calculate u_{xt}^{ij}. In other words, I can not directly estimate equation (6.3). Instead, I treat the choice probability as the model prediction for the market share and aggregate s_{ft}^{ij} over j to calculate the unconditional market share.

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

$$s_{ft}^{i} = \sum_{j} a_{t}^{ij} s_{ft}^{ij} \tag{6.4}$$

Denote the market size as λ_{t}^{i}. We can calculate the potential demand before matching process as:

$$u_{ft}^{i} = \lambda_{t}^{i} s_{ft}^{i} \tag{6.5}$$

In comparison to the exogenous destination distribution of travellers A_{t}^{i}, I can also calculate the destination distribution of each firm $\hat{A}_{ft}^{i}$ as outcomes of passengers' discrete choice using Bayes' rule. Thus, the model predicted firm specific destination distribution becomes:

$$\hat{a}_{ft}^{ij} = \frac{a_{t}^{ij} s_{ft}^{ij}}{s_{ft}^{i}} \tag{6.6}$$

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

To summarize the demand side, I assume passengers make demand decision before matching process but with knowledge of matching probability and post-match utility. Given a set of demand parameter values, the demand model can predict two main things. First, the model predicts market shares s^i_{ft} and demand u^i_{ft}. Second, it predicts endogenous destination distribution of each firm. However, we cannot directly observe demand or supply in the data . In the next section, I discuss how to predict supply and link model predicted supply and demand 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 $\hat{a}^{ij}_{ft}$ which is obtained from (6.6). 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 go 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 in the next period given N drivers and I locations will be C^{I-1}_{N+1} which is large when N is large[20]. 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 shock by choosing the location with maximum value:

[20]This number is obtained by counting the number of outcomes of putting N balls in I urns $C^{I-1}_{N+1} = \frac{(N+1)!}{(I-1)!(N-I+2)!}$.

$$j^* = \arg\max_j \{V^j_{ft+\chi^{ij}_t} - c^{ij}_t + \epsilon^j\} \tag{6.7}$$

where c^{ij}_t is the cost of travelling to j (gas), ϵ^j is supply shock and $V^j_{ft+\chi^{ij}_t}$ is driver's ex-ante value of searching location j in period $t + \chi^{ij}_t$. The number of periods travelling from i to j at t is χ^{ij}_t. In other words, χ^{ij}_t is the time cost from i to j. When the driver chooses j which is far from i, he has to account for the loss of not searching for passengers in 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 time cost and distance cost are allowed to vary over t. The ex-ante value is defined as:

$$V^j_{ft} = \phi^j_{ft}\big(\sum_l \hat{a}^{jl}_{ft}(p^{jl}_{ft} - c^{jl}_t + V^l_{ft+\chi^{jl}_t})\big) + (1-\phi^j_{ft})\mathbb{E}_\epsilon\big(-\mu + \max_l\{V^l_{ft+\chi^{jl}_t} - c^{jl}_t + \epsilon^l\}\big). \tag{6.8}$$

In equation (6.8), 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 $\hat{a}^{jl}_{ft}$. Recall that $\hat{a}^{jl}_{ft}$ measures firm specific destination distribution obtained in (6.6). 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 (6.8) is the continuation value of not being matched in j. Conditional on not being matched, a search cost μ occurs[21]. Since drivers do not observe realized supply shock until the end of period, the continuation value takes expectation over all possible supply shocks. I assume the supply shock ϵ follows i.i.d T1EV distribution with scale parameter σ such that the continuation value of unmatched has explicit form:

[21] This constant turn plays an important role to make values conditional on not being matched less preferable than values condition on being matched. Due to the existence of supply shock and large number of locations, $\mathbb{E}_\epsilon \max_l\{\epsilon^l\} = \sigma \log I$ could be larger than expected fares.

$$\mathbb{E}_{\varepsilon} \max_{l} \{V^l_{ft+\chi^{jl}_t} - c^{jl}_t + \epsilon^l\} = \sigma \log \left(\Sigma_l \exp((V^l_{ft+\chi^{jl}_t} - c^{jl}_t)/\sigma) \right) \tag{6.9}$$

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:

$$\pi^{ij}_{ft} = \frac{\exp((V^j_{ft+\chi^{ij}_t} - c^{ij}_t)/\sigma)}{\Sigma_l \exp((V^l_{ft+\chi^{il}_t} - c^{il}_t)/\sigma)} \tag{6.10}$$

The scale parameter σ controls for incentives of drivers searching certain locations captured by continuation values other than shocks. For example, large σ 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 movement of employed cars following the destination distribution of passengers $\{\hat{a}^{ij}_{ft}\}$. Combining the transition of employed cars $\hat{A}_{ft}$ and the policy function of unemployed cars Π_{ft} of equation (6.10) gives the law of motion of state transition. The state includes the status of all in-transit cars. The state at the beginning of period t is a collection of $\{S^i_t\}_{\forall i}$ where S^i_t is a collection of $\{\tilde{v}^i_{ft,k}\}_{f,k}$ with $\tilde{v}^i_{ft,k}$ indicating the number of cars from 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 $v^i_{ft} = \tilde{v}^i_{ft,k=1}$. At the end of each period, the transition of employed and unemployed cars update the state such that:

$$\tilde{v}^i_{ft+1,k} = \tilde{v}^i_{ft,k+1} + \Sigma_j m^j_{ft} \hat{a}^{ji}_{ft} \mathbb{1}_{\chi^{ji}_t = k} + \Sigma_j (v^j_{ft} - m^j_{ft}) \pi^{ji}_{ft} \mathbb{1}_{\chi^{ji}_t = k}, \forall f, i, k \tag{6.11}$$

To interpret (6.11), at beginning of period $t+1$, the number of firm f drivers that will arrive at location i 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 assume a matching 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 derived by Burdett, Shi and Wright (2001). I use the same functional form as Buchholz (2016) with a modification to reflect heterogeneity across locations. This matching process is only applied to taxis within a location. For Uber, I assume perfect matching within a location. I will discuss both functions below.

First, focus on the random matching of taxis. Given taxis' demand u^i_{yt} and supply v^i_{yt} 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. 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 $1-(1-1/v^i_{yt})^{u^i_{yt}}$. All taxis have the same probability of being matched, and therefore the number of matches is:

$$\begin{aligned} m(u^i_{yt}, v^i_{yt}) &= v^i_{yt}\Big(1-(1-\frac{1}{v^i_{yt}})^{u^i_{yt}}\Big) \\ &= v^i_{yt}(1-\exp(-\frac{u^i_{yt}}{v^i_{yt}})) \end{aligned} \tag{6.12}$$

Function (6.12) itself allows matching friction due to coordination failures such that there is possibility that some cars receive no visits. To reflect the heterogeneity of locations and parameterize the matching efficiency, I modify equation (6.12) such that:

$$m(u^i_{yt}, v^i_{yt}) = v^i_{yt}(1-\exp(-\frac{u^i_{yt}}{\gamma_1 v^i_{yt}+\gamma_2+\gamma_3 block^i})) \tag{6.13}$$

Basically, the new matching function allows location heterogeneity by adding wrong blocks that passengers could visit[22]. The larger the area of the location is, the more

[22] A block is defined as 0.2 square kilometer. The number of blocks in location i is calculated from dividing area size(square kilometers) of location i by 0.2.

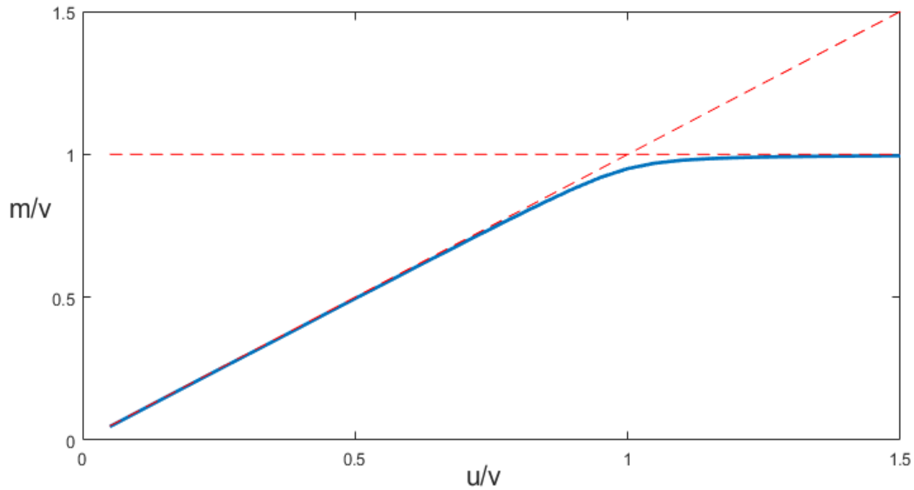


Figure 2: Hyperbolic function for perfect matching

wrong blocks that passengers may visit. Moreover, γ_1 measures the efficiency of taxis such that large value of γ_1 generates fewer matches as Buchholz (2016) shows.

As for Uber, the matching between Uber drivers and passengers is frictionless thanks to the mobile app matching technology. A passenger can observe the exact location of the driver once being matched and vice versa. In other words, once demand and supply are realized, matches can be perfectly made without the coordination failure present with taxis. 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 invertible. 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 2. 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 2 is obtained by solving m^i_{xt} from (with small value of ϵ):

$$(\frac{m^i_{xt}}{v^i_{xt}} - 1)(\frac{m^i_{xt}}{v^i_{xt}} - \frac{u^i_{xt}}{v^i_{xt}}) = \epsilon \tag{6.14}$$

The limitation of this matching function is that when demand is greater than supply, the potential demand is obtained solely by curvature of the hyperbolic function. 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 $\{\mathcal{S}_t^i\}_{i \in I}$. For any i, $\mathcal{S}_t^i$ includes information about arrival of cars in next K periods, hence collection $\{\tilde{v}_{ft,k}^i\}_{f,k \in K}$. In OE, drivers make optimal search decisions according to $\{s_t, \{\mathcal{S}_t^i\}_{\forall i,t}\}$. Drivers' belief on the evolution of market state (i.e. supply distribution) is consistent with the realized state in equilibrium[23]. 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 $\{v_{ft}^i\}$, beliefs of state transition $\{\hat{v}_{ft,k}^i\}$, policy function of unemployed cars $\{\pi_{ft}^{ij}\}$, transition of employed cars $\{\hat{a}_{ft}^{ij}\}$ for $\forall i, t, f$ and given initial distribution of supply $\{v_{ft=1}^i\}$ such that:

1. At the beginning of each period t, in any location i, passengers make discrete choice between firms based on (6.1)-(6.3). Market demand is calculated from (6.4-6.5).

2. Matches are made randomly between supply and demand for each firm, location and time. Within location, the matching process follows (6.13) for taxis and (6.14) for Uber.

3. Transition of employed cars follows $\{\hat{a}_{ft}^{ij}\}$ obtained by Bayes' rule (6.6).

4. At the end of each period, unemployed drivers follow policy function π_{ft}^{ij} calculated in (6.10) based on beliefs of state transition $\{\hat{v}_{ft+1,k}^i\}$.

5. Realized state transition is obtained by combining both employed ($\{\hat{a}_{ft}^{ij}\}$) and unemployed cars ($\{\pi_{ft}^{ij}\}$). State of next period is updated to $\tilde{v}_{ft+1,k}^i$ by (6.11).

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

[23]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 $\{V_{ft}^i\}_{\forall i,t}$. Knowing supply distribution or search values are interchangeable given one step calculation of (6.8).

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 location-time 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 $\{\delta^i_{ft}\}_{\forall f,t,i}$ and price coefficients $\{\alpha_1, \alpha_2\}$. The supply side parameters include search cost μ and supply shock parameter σ. Finally, there are parameters defining the matching function of taxis $\{\gamma_1, \gamma_2, \gamma_3\}$. This process can be summarized as figure below:

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 destination distribution of taxis in the data.

In general, the process is similar to estimating a random coefficient discrete choice model with complexities. Given a set of nonlinear parameter values $\{\alpha, \gamma, \sigma\}$, we solve the fixed point for mean utilities $\{\delta^i_{ft}\}$ such that the model predicted shares of pickups equal shares in the data[24]. Given $\{\delta^i_{ft}\}$, I estimate nonlinear parameters using nonlinear least square estimation. Parameters within δ are estimated using linear

[24]In Nevo (2000), he denotes parameters other than δ as nonlinear parameters since they nonlinearly enter GMM objective function and parameters within δ as linear parameters.

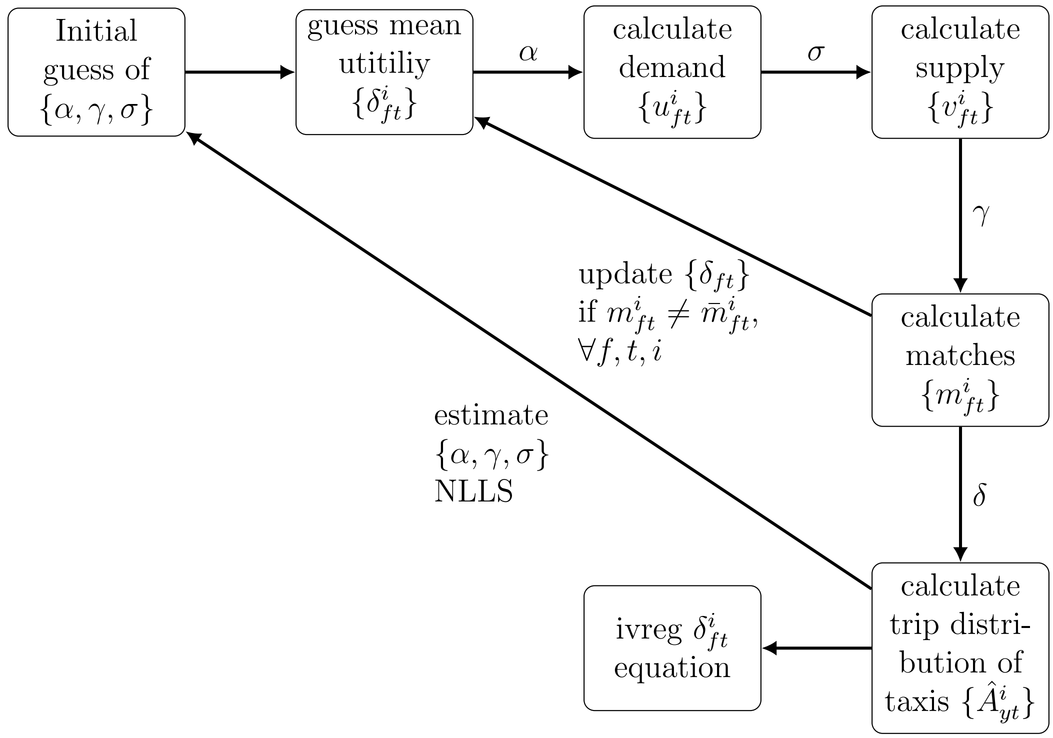


Figure 3: overview of the estimation process

IV regression because of endogeneity of supply in demand equation. The instrument variable I use for supply is dropoffs of cars in the market. 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 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. For taxis, Buchholz (2016) uses the number of medallions as total taxis in his model. In Frechette et al.(2016), they find that almost 80% of minifleets and 70% of self-owned taxi medallions are

			$t+1$				
Type of Driver	% of Active Drivers	t	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 4: Uber drivers' working schedules from Chen et. al.(2017)

active during day shift hours. Thus, I assume the fixed number of taxis in my model as 13,587. 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 4. 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 4. In the stationary distribution, there are 10% morning drivers which is equivalent to 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 is 3,000.

7.1.2 Passengers' Destinations

As mentioned in the sample construction, in the demand side I need to know the exogenous trip 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: the pattern doesn't change due to Uber's entry and 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 origins can be nonparametrically calculated using taxis trip records from 2010.

7.2 Two-Step Estimation

7.2.1 Step 1: Estimating Nonlinear Parameters

The first step is to estimate $\{\alpha, \sigma, \gamma, \delta^i_{ft}\}$. The estimation process can be further broken into two procedures. First, given the parameter values of $\{\alpha, \sigma, \gamma\}$, we need to solve equilibrium demand and supply subject to $\{\delta^i_{ft}\}$ such that model generates the same pickups as in the data. Second, in the outer loop of first procedure, we need to update $\{\alpha, \sigma, \gamma\}$ to minimize deviation of model predicted trip patterns of employed taxis to the data. The first procedure follows Buchholz (2016) with modification to discrete choice demand. In Buchholz (2016), he searches for equilibrium demand $\{u^i_{ft}\}$ which generate pickups in the sample. Instead of solving for demand $\{u^i_{ft}\}$, I am solving for fixed mean utilities $\{\delta^i_{ft}\}$ which have one-to-one mapping to demand according to BLP contraction mapping. Solving mean utilities is necessary to calculate trips of passengers choosing specific firm $\{\hat{A}^i_{ft}\}$ based on what drivers' search values $\{V^i_{ft}\}$ are calculated.

The first procedure is as follows. Given parameter values $\{\alpha, \sigma, \gamma\}$, I make initial guess of all mean utilities $\{\delta^i_{ft}\}$ which are obtained by treating pickups as demands and BLP contraction mapping with demand shares. Given the mean utilities and demands, we need to figure out corresponding equilibrium supplies $\{v^i_{ft}\}$. Instead of an arbitrary guess of supplies, I use backward induction to calculate my starting point for fixed point iteration of equilibrium supply. From the last period T, given the demand distribution and zero search values $V^i_{ft} = 0, \forall t > T$, I can calculate continuation values $\{V^i_{fT}\}$ for an arbitrary supply distribution $\{v^i_{fT}\}$[25]. 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 $\{v^i_{ft}\}$. 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

[25] 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 (6.10).

Algorithm 1 Solve Equilibrium Supply

```
1: Set parameter values for {σ, α, γ} and {δ^i_{ft}}_{∀f,i,t}
2: Guess supply {v^i_{fT}}_{f,i}, calculate {u^i_{fT}}_{f,i}, {V^i_{fT}}_{f,i}
3: for τ = T − 1 to 1
4:     Guess {v^i_{fτ}}_{f,i}, reset state of in-transit cars {ṽ^i_{ft,k}}_{t>τ} = 0
5:     for t = τ 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 {â^i_{ft}}
9:         Compute transition policy of unemployed cars {π^i_{ft}}
10:        Update {ṽ^i_{ft+1,k}} based on {π^i_{ft}}, {â^i_{ft}}
11:        Update value {V^i_{ft}} of current period
12:        Update supply of next period v^i_{ft+1} = ṽ^i_{ft+1,k=1}
13:    end
14:end
15:Fix τ = 1
16:    Iterate step 5 to 13
17:    Stop until step 11 and 12 won't update under certain tolerance level.
```

1 below. I find starting point of supplies for iteration in step 2-14. Step 15-17 is the iteration for equilibrium supplies[26]. This procedure generates mapping from mean utilities(or demands) to supplies, denoted as $\mathbf{v} = \Gamma(\boldsymbol{\delta})$. 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 $\{\bar{m}^i_{ft}\}$, 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 $\{\alpha, \sigma, \gamma\}$. These two algorithms produce the transition of employed taxis in the end. 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 7 larger areas over 10 hours. This process applies to both data and model generated outcomes. Then I calculate the sum of squared deviations between model generated transition of taxis to the data as my nonlinear least squares objective function for estimation.

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

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 $\{\delta^i_{ft}\}_2$ to match shares
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 $u^i_{xt} = \bar{m}^i_{xt}$ 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 $\{\hat{A}^i_{yt}\}$, transition of employed taxis

7.2.2 Step 2: Estimating Demand Parameters

Given the estimates of $\{\alpha, m^i_{ft}, \delta^i_{ft}, u^i_{ft}, v^i_{ft}\}$ in step 1. I can calculate some variables in mean utility such as demand side's matching probability and demand-to-supply ratio (proxy for waiting time) of Uber as in equation (6.2). In the specification of mean utility, the coefficient on matching probability equals 1 by model assumption. The endogeneity issue occurs for estimating coefficient on demand-to-supply ratio of Uber. Demand itself is function of unobserved demand shock ξ^i_{ft}. The endogenous supply also depends on demand shock ξ^i_{ft} when unemployed drivers make search decisions. In order to solve for the issue, I use arrival of employed cars as instrument for supply. The validity of this instrument is as follows. First, this instrument is correlated with supply since it is part of supply together with arrival of unemployed cars. Second, it is uncorrelated with demand shocks since its arrival is subject to the destinations of passengers which are exogenous.

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 riderships 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 (6.6) 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 (6.6). 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[27]. 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 on demand shocks, I need to find instrument for demand-to-supply ratio of Uber (proxy for waiting time)[28]. I choose arrival of employed taxis as instrument which is correlated with supply and thus demand-to-supply ratio and is uncorrelated with demand shocks. The estimation result is in next section.

8 Results

The estimation results are listed in table 4 below. The estimates of price coefficients for different trip distance are $\alpha_1 \exp(-\alpha_2 distance)$ as in utility form (6.2). The positive sign of α_2 indicates that the price elasticity is smaller for those travelling a long distance. The supply shock parameter σ in equation (6.10) is 15 in comparison to 12.5 estimated in Buchholz (2016). However, the NLLS objective function is not very sensitive to this parameter. This parameter σ plays a similar role as a discount factor which affects search values $\{V_{ft}^i\}$ uniformly. Thus, it may be hard to identify it. The parameters of the matching functions in equation (6.13) are $\{\gamma\}$ in which γ_1 is the coefficient on supply and γ_2, γ_3 measure location heterogeneity in area sizes. A large γ_1 means there are $(\gamma_1 - 1)v_{ft}^i$ location spots greater than actual number of taxis that passengers may randomly go to and therefore the matching efficiency decreases. For each location i divided into blocks, the extra spots that passengers may search are $\gamma_2 + \gamma_3 block^i$ which implies that the within-location matching efficiency of taxis is high when values for γ_2, γ_3 are small. Parameter μ as in equation (6.8) measures the fixed cost for each search which is \$0.5. This parameter μ does not affect the search decisions of unemployed cars since it is cancelled out in policy function (6.10). Instead, it measures the difference between the expected profit of being matched and the expected profit of not being matched. It also shifts the ex-ante search values $\{V_{ft}^i\}$.

Finally, there are 4800 mean utilities to estimate and the statistics for taxi or

[27] For a given market i, t, the mean utility δ_{ft}^i 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.

[28] The matching probability is also endogenous. However, by the utility specification, coefficient on logarithm of matching probability equals 1.

Uber is listed in the bottom of table 4. The mean of taxis' $\{\delta^i_{yt}\}$ is 7.37 with maximum value at 12.19 and minimum value at 2.14. Uber's mean utilities $\{\delta^i_{xt}\}$ are less than taxis' after controlling for prices. One explanation is that Uber's supply is far less than taxis and passengers may have long waiting time to get a Uber which makes the mean utility of choosing Uber small[29]. The second explanation is that not all passengers have installed Uber app or even know about Uber. This effect is captured in Uber's dummy variable such that d_x is smaller than taxi fixed effect which in turn makes the estimated δ^i_{xt}, on average, smaller than δ^i_{yt}. I am still working on bootstrap for standard errors of the estimates.

Table 4: Estimates of nonlinear parameters

panel 1: nonlinear parameter		*estiamtes*
α_1		-3.94
α_2		0.08
σ		15
γ_1		1.2
γ_2		10
γ_3		1
μ		0.5
	mean	*min/max*
δ^i_{yt}	7.37	2.14/12.19
δ^i_{xt}	5.89	1.07/10.73

The statistics of three variables in equilibrium including demand, supply, and search values are shown in table 5 separately for taxi and Uber. The mean of taxis' demand in 2400 location-period is 136 with maximum demand of 809.5 and minimum of 1.97. In comparison to taxis, Uber's potential demand is less with 20.31 on average. The average supply of taxis for each location-period is 150 ranging from 1.78 to 1779.3. The average supply of Uber is 33. Finally, the ex-ante search values for taxis have the maximum value of \$214 at the beginning of the day (6 a.m.). Uber's search values are \$281.24 at the beginning of the day and average \$130.67. Both are higher than the values for taxis. This high profitability of Uber could be the result of surge pricing, matching efficiency and less cannibalization. The minimum fare \$7 and surge multiplier make the expected profit of Uber conditional on being matched higher than taxis.

[29]Uber's matching probability is higher than taxis since my estimates of Uber's equilibrium supply is greater than demand/pickups for most markets. But there is demand-to-supply ratio of Uber (see (6.2)) as proxy for waiting time which may make δ^i_{xt} less than δ^i_{yt} as well as Uber dummy variable d_x.

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: Statistics in equilibrium

demand	*mean*	*min/max*
u^i_{yt}	136.73	1.97/809.50
u^i_{xt}	20.31	1.06/51.72
supply		
v^i_{yt}	150	1.78/1779.3
v^i_{xt}	33	4.05/227.7
search values		
V^i_{yt}	91.87	0.86/214.32
V^i_{xt}	130.67	5.36/281.24

Given the estimates in step 1, I run linear regression of mean utilities $\{\delta^i_{ft}\}$ on variables in equation (6.2). I rewrite the equation such that the dependent variable is $\delta^i_{ft} - \log(\tau^i_{ft})$ where τ^i_{ft} is calculated with equilibrium demand u^i_{ft} and supply v^i_{ft}. With the linear estimates in (6.2), I can calculate new $\{\delta^i_{ft}\}$ in couterfactuals. The OLS and IV regression results are given in table 6. Parameter θ is the coefficient on Uber's demand-to-supply ratio. Since both demand and supply depend on the demand shock ξ^i_{ft}, this variable is endogenous. The demand-to-supply ratio is positively correlated with demand shock if positive influence of ξ^i_{ft} on demand u^i_{ft} dominates positive influence on supply v^i_{ft}. Otherwise, the correlation is negative. In OLS regression, $\hat{\theta}$ is 0.83 which implies that a high demand-to-supply ratio increases the demand after controlling for matching probability τ^i_{ft}. Conditional on being matched, I use the demand-to-supply ratio as a proxy for waiting time and the positive coefficient is counter intuitive. Using an instrument variable, which is the arrival of employed cars, the sign of θ becomes negative implying high demand-to-supply ratio increases waiting time and therefore decreases mean utility as expected. The validity of such instrument variable is argued in estimation section. First, arrival of employed cars are determined by destination of passengers on board which is uncorrelated with demand shock in the destination. Second, arrived cars is part of supply in the destination of current period such that it is correlated with the endogenous variable instrumented. Uber's fixed effect is -2.69 which means that on average Uber is less attractive to passengers after controlling for prices, matching probability and waiting time. The reason of this small

Uber fixed effect as discussed before could be that not all passengers would consider Uber because of the cost of installing mobile app or because of safety concerns.

Table 6: Linear regression of mean utility

Dependent variable $\delta^i_{ft} - \log(\tau^i_{ft})$		
	OLS	*IV*
θ	0.83	-1.23
d_x	-1.57	-2.69
constant	6.32	4
location fixed effects	YES	YES
time fixed effects	YES	YES

Figure 5-7 shows the intra-location trend of dynamic equilibrium supply, demand, and matches over 60 time periods, from both the data and model, for three selected locations: Time Square, PennStation and JFK airport. The top picture in each figure is for taxis and bottom is for Uber. As assumed in the model, taxis have mismatches within location due to the random matching process, whereas Uber has perfect matching. To reflect in the figure, the within location mismatches of taxis and passengers can be observed in first picture of figure 5 that both supply curve v and demand curve u are above the curve for matches. There are two curves for matches, one is generated by the model and one comes from the data. In the estimation, I fit model predicted pickups to the data which results in the overlap of these two curves.

In Time Square, the supply of taxis are high in the morning and low in the afternoon compared to demand which is growing over time. In the morning, there is excess supply in Time Square and demand becomes greater than supply in the afternoon. The gap between match curve and lower bound of supply and demand curve is mismatches due to randomness for each fixed time period. Unlike taxis, the supply curve of Uber in Time Square is above demand/match curves. However, Uber's supply in Time Squire is around 60 vehicles far less than taxis which is more than 200 over periods. During the morning when Time Square exhibits excess supply, there is excess demand in Penn Station which is near to Time Square as shown in figure 6. Figure 6 shows that the supply of taxis in Penn Station is lower than demand throughout the whole daytime shift. Taxis' supply and demand in figure 5&6 show the cross location mismatches that one location has excess supply and one has excess demand simultaneously.

At the airports, both Uber and taxis have frictionless matching process such that $m = \min\{u, v\}$. Figure 7 shows the dynamics in JFK airport. There is not within

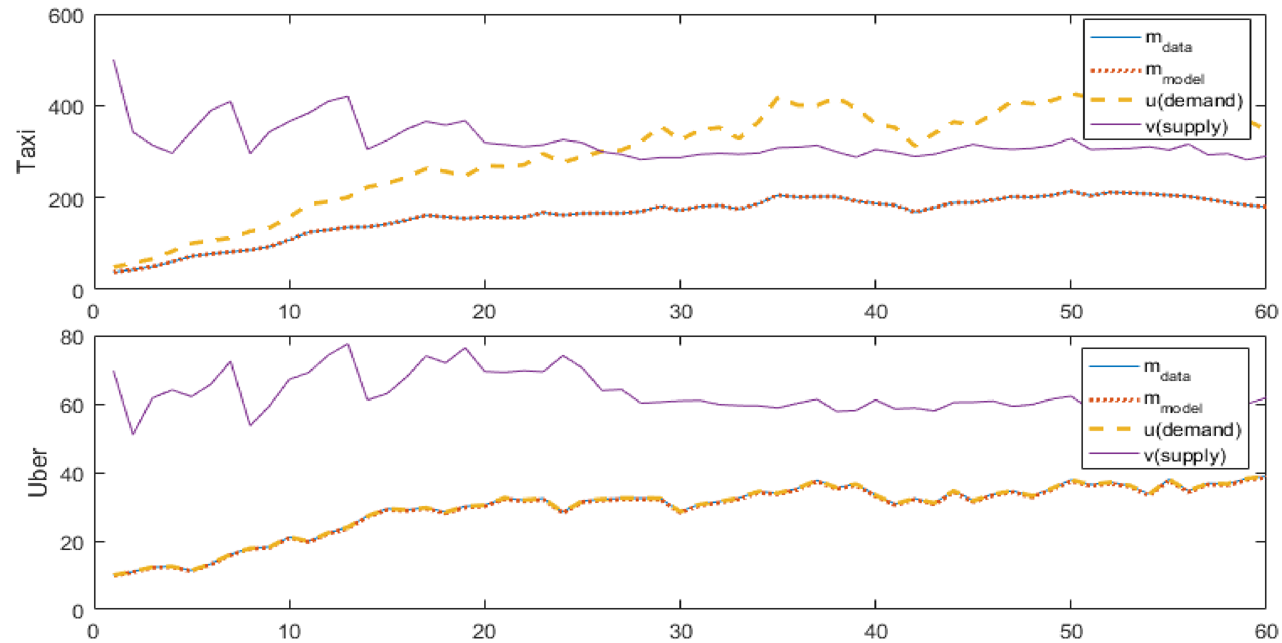


Figure 5: Taxi&Uber supply/demand/match over time in Time Square

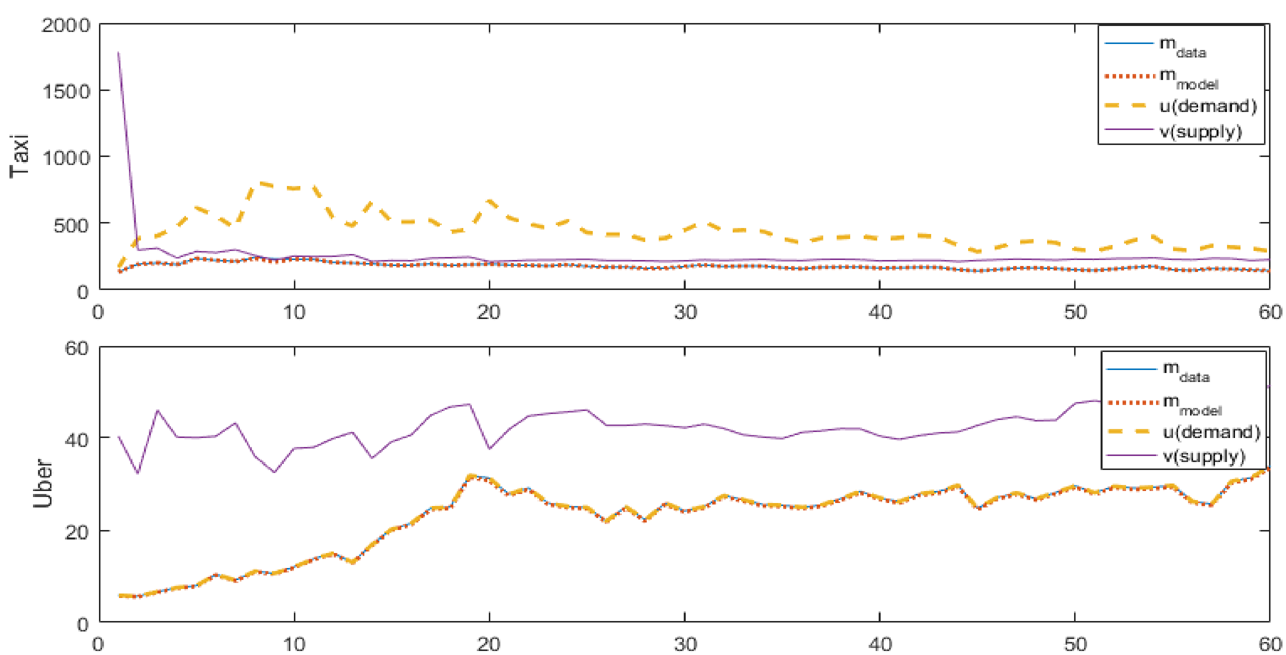


Figure 6: Taxi&Uber supply/demand/match over time in Penn Station

location mismatches for both taxis and Uber cars. The estimated supplies of taxis and Uber are larger than demand in equilibrium. In all three selected markets and many other markets not shown here, Uber has excess supply which could result from the large number of total Uber cars taht is assumed to be 3000. Due to the limitation of my data, I can not tell whether Uber's supplies are overestimated or not. Moreover, the aggregate number of Uber cars may also vary over time which is not modelled in this paper. Under current estimates, the matching probabilities for passengers choosing Uber are almost one. Thus, in the counterfactual of restricting Uber's supply of next section, only demand-to-supply ratio (waiting time) for Uber will respond to the policy change leaving matching probability unchanged.

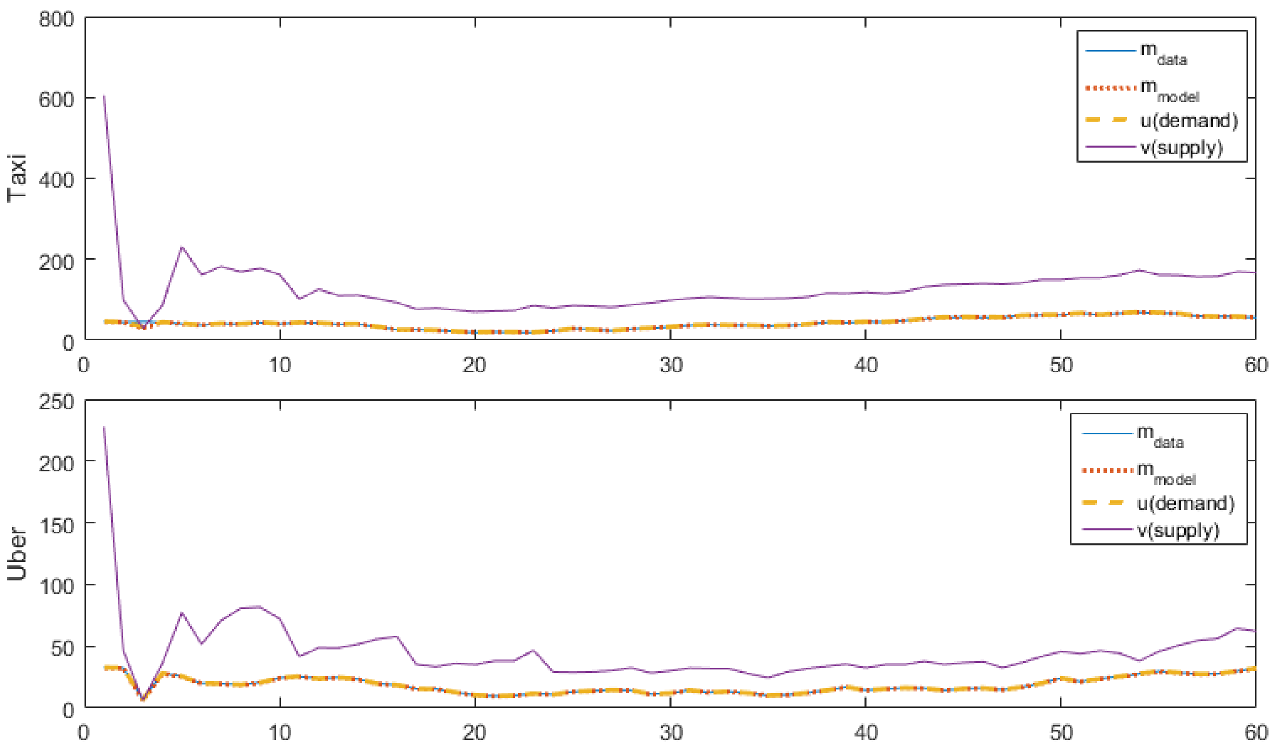


Figure 7: Taxi&Uber supply/demand/match over time in JFK

The baseline welfare statistics, firms' profits, consumer welfare, and within/cross location mismatches, are summarized in table 7. The within-location mismatches count matches that could be made within market without coordination failure. Cross-location mismatches count matches that could be made by efficient allocation of drivers within periods. Consumer welfare is calculated as inclusive value of discrete choice

$\log(\Sigma_f \exp(\delta^i_{ft} + \alpha^{ij}_{ft} p^{ij}_{ft}))$ prior to matching process[30]. The over time trend of these statistics is depicted in figure 8.

Table 7: Baseline Welfare Statistics (daily aggregate)

<u>within-location mismatches</u>		
$\Sigma_{i,t} \min\{u^i_{yt}, v^i_{yt}\} - \bar{m}^i_{yt}$		94,454
$\Sigma_{i,t} \min\{u^i_{xt}, v^i_{xt}\} - \bar{m}^i_{xt}$		680
<u>cross-location mismatches</u>		
$\Sigma_t \min\{\Sigma_i \max\{u^i_{yt} - v^i_{yt}, 0\}, \Sigma_i \max\{v^i_{yt} - u^i_{yt}, 0\}\}$		49,298
$\Sigma_t \min\{\Sigma_i \max\{u^i_{xt} - v^i_{xt}, 0\}, \Sigma_i \max\{v^i_{xt} - u^i_{xt}, 0\}\}$		860.45
<u>within-location mismatches</u>		
taxi profit		$ 2,549,100
Uber profit		$ 827,450
consumer welfare		734,860
<u>matches in data v.s. model generated matches</u>	data	model
$\Sigma_{i,t} m^i_{yt}$	173,490	170,990
$\Sigma_{i,t} m^i_{xt}$	48,897	47,198

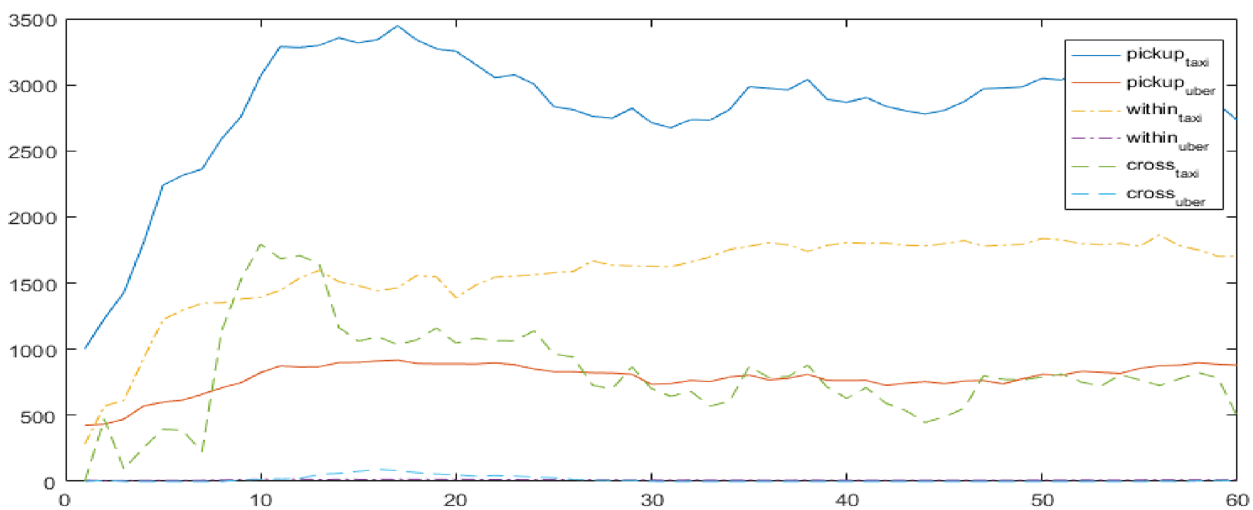


Figure 8: within/cross location (mis)matches of taxis& Uber over time

[30]According to (6.2) this inclusive value evaluates logarithm of indirect utility.

9 Counterfactuals

In this section, I simulate three counterfactual scenarios preserving the demand shocks $\{\xi_{ft}\}$ and parameter estimates. The first counterfactual is regulating Uber' supply corresponding to NYC government's proposal to cap growth of Uber in summer 2015. The second counterfactual is improving traffic conditions of the city. Though government blames Uber for causing traffic congestion, I do not endogenously model traffic condition in my model. Instead, the counterfactual intends to show how traffic condition affects matching efficiency of the industry. The last counterfactual is to test the effectiveness of Uber's surge multiplier on matching efficiency of Uber drivers. Uber uses flexible pricing in contrast to taxi fixed fare to coordinate supply and demand. I would like to test how effective it is under this search and matching model. In all three counterfactuals, I disentangle the effect of indirect network effect by shutting down the response of demand to changes of supply at first and allow full response between demand and supply in the end. To be specific, I shut down response of demand to changes of equilibrium supply by fixing mean utility $\{\delta^i_{ft}\}$ such that matching probability and demand-to-supply ratio within $\{\delta^i_{ft}\}$ do not react to changes of supply which mimics previous work of Buchholz (2016) without network effects.

9.1 Regulating Uber's supply

In June 2015, city government proposed to cap the growth of Uber. At the time, the number of Uber affiliated cars is growing at monthly rate 3%. The government proposes to restrict the growth to 1% per year. Assuming Uber keeps growing at 3% and 10 months later in April 2016, the total number of Uber increases by 30%. Supposing the proposal is implemented immediately in June 2015, the total supply of Uber will be 30% less. Thus, to test the effect of the policy, I roughly discount the total number of Uber drivers in my model by 1.3. Thus the total number of Uber cars becomes 3000/1.3 and I solve for the new dynamic equilibrium using the search and matching model. I solve equilibriums with or without indirect network effects. The statistical summary of new equilibriums compared to baseline is provided in table 8.

The first column of table 8 is statistics of the baseline case. In the second column, equilibrium is solved without response of demand to supply by fixing $\{\delta^i_{ft}\}$. The last column is new equilibrium in full model. In table 8, taxis' supply, demand and pickups do not change significantly under supply regulation of Uber. Taxis' pickups increase slightly from 170,990 to 171,040 without network effect and to 172,030 with network effects. The demand and supply of Uber decrease significantly. For example, Uber's aggregate daily supply decreases from 78,927 to 61,347 and demand decreases from 48,739 to 42,094. From table 8, we can tell that most of Uber's past passengers choose subway after supply regulation instead of switching to taxis. The cross-location mismatches for Uber increase from 860 to 4037 without considering demand's response to supply changes. However, with network effects the cross location mismatches are less which indicates the importance of considering network effects in policy application.

Table 8: Uber Supply Regultion

two type friction	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t}v^i_{yt}$	360,440	360,610	360,630
$\Sigma_{i,t}v^i_{xt}$	78,972	62,223	61,347
$\Sigma_{i,t}u^i_{yt}$	328,140	328,090	330,870
$\Sigma_{i,t}u^i_{xt}$	48,739	48,735	42,094
$\Sigma_{i,t}m^i_{yt}$	170,990	171,040	172,030
$\Sigma_{i,t}m^i_{xt}$	47,198	43,930	40,158
$\Sigma_{i,t}\min\{u^i_{yt}, v^i_{yt}\} - \bar{m}^i_{yt}$	94,454	94,554	94,621
$\Sigma_{i,t}\min\{u^i_{xt}, v^i_{xt}\} - \bar{m}^i_{xt}$	680	729	636
$\Sigma_t \min\{\Sigma_i \max\{u^i_{yt} - v^i_{yt}, 0\}, \Sigma_i \max\{v^i_{yt} - u^i_{yt}, 0\}\}$	49,298	49,210	49,529
$\Sigma_t \min\{\Sigma_i \max\{u^i_{xt} - v^i_{xt}, 0\}, \Sigma_i \max\{v^i_{xt} - u^i_{xt}, 0\}\}$	860	4037	1299
$\$taxiprofit$	$2.55 * 10^6$	$2.55 * 10^6$	$2.57 * 10^6$
$\$Uberprofit$	$0.82 * 10^6$	$0.76 * 10^6$	$0.70 * 10^6$
$consumer\ welfare$	$7.35 * 10^5$	$7.35 * 10^5$	$7.26 * 10^5$

Another important observation in this counterfactual is that the profit of taxis does not increase significantly after regulating Uber since most Uber's passengers switch to subways in new equilibrium. This finding suggests that Uber may not "steal" rent directly from taxis by competing for passengers. The data indicates a very different supply distribution between Uber and taxis that Uber serves outer boroughs more than taxis. Rather than competition, loss of taxis' rents may result from worse traffic conditions in NYC such that it takes longer time for delivering trips or searching across locations. Unless government has evidence that Uber contributes to traffic congestion, capping Uber may not increase taxis' rents. At last, consumer welfare is worse off under Uber regulation. In order to understand influence of traffic conditions, in the next subsection I simulate new equilibrium after improving traffic.

9.2 Improving traffic conditions

In this section, I simulate equilibrium after improving traffic conditions. To do so, I assume the travelling time $\{\chi_t^{ij}\}$ from location i to j in period t is the same to that in 2010 before Uber's entry[31]. The summary statistics of new equilibrium without and with network effects are listed in table 9.

When traffic conditions improve, the daily aggregate supply of both Uber and taxis increase. For example, taxis' supply increase from 360,440 to 409,820. The demand for both taxis and Uber increase slightly since increased supply shifts mean utilities of choosing taxis or Uber upwards. Total pickups of taxis increase from 170,990 to 181,080 and pickups of Uber increase from 47,198 to 50,758. As for within location friction, taxis' mismatches increase due to increase of both demand and supply. This measure counts the absolute number of mismatches rather than ratio of mismatches to demand or supply. Thus, more supply and demand results in more mismatches. The change of cross location mismatches are interesting which decrease in new equilibrium. This is due to the improved mobility of cars such that matching is more efficient compared to baseline case. However, comparing equilibrium with and without network effects, both frictions are underestimated without network effects.

Table 9: Traffic Improvement

two type friction	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v_{yt}^i$	360,440	409,570	409,820
$\Sigma_{i,t} v_{xt}^i$	78,972	89,130	89,798
$\Sigma_{i,t} u_{yt}^i$	328,140	328,090	331,030
$\Sigma_{i,t} u_{xt}^i$	48,739	48,753	51,953
$\Sigma_{i,t} m_{yt}^i$	170,990	180,000	181,080
$\Sigma_{i,t} m_{xt}^i$	47,198	47,851	50,758
$\Sigma_{i,t} \min\{u_{yt}^i, v_{yt}^i\} - \bar{m}_{yt}^i$	94,454	105,250	105,780
$\Sigma_{i,t} \min\{u_{xt}^i, v_{xt}^i\} - \bar{m}_{xt}^i$	680	656	773
$\Sigma_t \min\{\Sigma_i \max\{u_{yt}^i - v_{yt}^i, 0\}, \Sigma_i \max\{v_{yt}^i - u_{yt}^i, 0\}\}$	49,298	39,878	40,857
$\Sigma_t \min\{\Sigma_i \max\{u_{xt}^i - v_{xt}^i, 0\}, \Sigma_i \max\{v_{xt}^i - u_{xt}^i, 0\}\}$	860	227	421
$\$taxiprofit$	$2.55 * 10^6$	$2.67 * 10^6$	$2.68 * 10^6$
$\$Uberprofit$	$0.82 * 10^6$	$0.84 * 10^6$	$0.89 * 10^6$
$consumer\ welfare$	$7.35 * 10^5$	$7.35 * 10^5$	$7.53 * 10^5$

In terms of welfare, both taxi's and Uber's profits increase. Passengers are also better off. This exercise suggests great benefits to the industry if traffic is better

[31] The trip time for any route ij at t of a representative weekday of 2010 can be calculated in the same way as 2016. One thing to note is that $\chi_{t,2016}^{ij}$ is not necessarily larger than $\chi_{t,2010}^{ij}$ for $\forall i, j, t$

especially for taxis' profits. Table 8&9 suggest that loss of taxis' rents mainly results from worse traffic conditions in comparison to Uber's competition. If Uber contributes to the traffic congestion, then the net effects of capping Uber may have positive social welfare by combing these two counterfactuals[32].

9.3 Eliminating surge multiplier

In the third counterfactual, I am interested in whether Uber's surge multiplier affects matching efficiency of the market. I eliminate Uber's surge multiplier by setting all Uber's prices $\{p_{xt}^{ij}\}$ to normal ones calculated solely based on trip distance and time. By decreasing Uber's prices, Uber drivers' search values and conditional expected profit on being matched in certain markets decrease. For example, markets with high profitability due to high surge multiplier becomes less attractive to Uber drivers which reallocates Uber's equilibrium supply. Taxis are also affected in new equilibrium due to endogenous demand and supply competition. The simulation results are provided in table 10. The most direct effects of eliminating Uber's supply multiplier is that demand for Uber increase from 48,739 to 58,957. In the contrary, low price of Uber makes taxis' demand decrease from 328,140 to 323,110. In the end, Uber's pickups increase while taxis' pickups decrease.

Table 10: Eliminating surge multiplier

two type friction	*Benchmark*	*w/o network*	*with network*
$\Sigma_{i,t} v_{yt}^i$	360,440	358,260	359,460
$\Sigma_{i,t} v_{xt}^i$	78,972	82,790	81,802
$\Sigma_{i,t} u_{yt}^i$	328,140	312,460	323,110
$\Sigma_{i,t} u_{xt}^i$	48,739	77,986	58,957
$\Sigma_{i,t} m_{yt}^i$	170,990	165,600	169,250
$\Sigma_{i,t} m_{xt}^i$	47,198	57,487	53,783
$\Sigma_{i,t} \min\{u_{yt}^i, v_{yt}^i\} - \bar{m}_{yt}^i$	94,454	94,097	94,304
$\Sigma_{i,t} \min\{u_{xt}^i, v_{xt}^i\} - \bar{m}_{xt}^i$	680	792	829
$\Sigma_t \min\{\Sigma_i \max\{u_{yt}^i - v_{yt}^i, 0\}, \Sigma_i \max\{v_{yt}^i - u_{yt}^i, 0\}\}$	49,298	47,657	48,880
$\Sigma_t \min\{\Sigma_i \max\{u_{xt}^i - v_{xt}^i, 0\}, \Sigma_i \max\{v_{xt}^i - u_{xt}^i, 0\}\}$	860	6135	3505
$\$taxiprofit$	$2.55 * 10^6$	$2.48 * 10^6$	$2.53 * 10^6$
$\$Uberprofit$	$0.82 * 10^6$	$0.81 * 10^6$	$0.76 * 10^6$
$consumer\ welfare$	$7.35 * 10^5$	$7.69 * 10^5$	$7.47 * 10^5$

[32] For now, I simulate these two counterfactuals separately. To link these two counterfactuals, I need to model endogenous traffic conditions which is challenging since real-time data on vehicles other than taxi and Uber is not available.

As for matching frictions, it is interesting to see that cross location mismatches of Uber increase a lot from 860 to 3505 after eliminating surge multiplier. It indicates Uber's surge multiplier helps with efficient matching. The goal of surge multiplier is to clear the market by making locations with excess demand more profitable in order to attract drivers. The differences with or without network effects are also obvious. For example, without network effect, Uber's cross location mismatches are much greater than mismatches with network effects. At last, the profits of Uber decrease due to price drop which dominates increase of Uber's pickups. Consumer welfare increases thanks to the low price of Uber's ride.

10 Conclusion

In this paper, I study the role of network effects in matching friction with an application to taxi and Uber drivers searching for passengers in New York City. Due to the fixed pricing structure of taxis, market is not cleared in prices leaving spatial mismatches across geographic locations such that some areas have waiting passengers (excess demand) and some have vacant cars (excess supply). I show that network effect affects spacial mismatches in equilibrium and if ignored would bias the policy simulation results. In the paper, I model drivers' dynamic spacial search problems taking into account demand distribution across locations and passengers' static discrete choice demand taking into account supply level as proxy for matching probability or waiting time. This model not only incorporates network effect but also supply competition between taxis and Uber that causes discussions on regulating Uber. My model is suitable to analyze such regulation(i.e. cap number of Uber cars) since it allows responses of matching process and demand to changes of equilibrium supply. I simulate three counterfactual scenarios with the model. The first decreases Uber supply by 30%, the second improves traffic conditions and the third eliminates the Uber surge multiplier. I find that taxis' pickups increase by 5.9% if traffic improves but do not increase significantly under supply regulation. Taxis' profits increase by 1% under supply regulation and increase by 5.18% if traffic improves. Uber's search friction increases after eliminating the surge multiplier or restricting supply. Consumer welfare decreases if Uber supply is restricted. Without network effects, search frictions and pickups will be underestimated.

References

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

Blanchard, O. and P. Dianmond (1989):“The Beveridge Curve,” *Brookings Papers on Economic Activity*, 1989, 1-76.

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

Buchholz, N.(2016):“Spatial Equilibrium, Search Frictions and Efficient Regulation in the Taxi Industry,” working paper.

Butters, G. R. (1977):“Equilibrium distributions of sales and advertising prices” *Review of Economic Studies*, 465-491.

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

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

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

Cullen, Z. and C. Farronato(2016):“Outsourcing Tasks Online:Matching Supply and Demand on Peer-to-Peer Internet Platforms”,working paper.

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.

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

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

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.

Gowrisankaran, G., M. Park, and M. Rysman (2014):“Measuring Network Effects in a Dynamic Environment,” *working paper*

Hopenhayn, H.A.(1992):“Entry, exit, and firm dynamics in long run equilibrium,” *Econometrica*,1127-1150.

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

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

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

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.

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.

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

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

Petrongolo, B. and C. A. Pissarides (2001):“ Looking into the Black Box: A Survey of the Matching Function,” *Journal of Economic Literature*, 390-431.

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

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

Shy, O. (2011): " A Short Survey of Network Economics," *Review of Industrial Organization*, 38,119-149.

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

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

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

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):

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

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:

$$v_{y1}^* = \frac{p_{y1}}{p_{y2}}u_{y1}^* \tag{T1}$$
$$u_{y1}^* = \frac{D_{y1}}{1+\alpha-\beta\dfrac{p_{y1}}{p_{y2}}} \tag{T2}$$
$$v_{y2}^* = N_y - v_{y1}^* \tag{T3}$$
$$u_{y2}^* = \frac{\beta}{1+\alpha}v_{y2}^* + \frac{D_{y2}}{1+\alpha} \tag{T4}$$

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 $u^*_{xi} < v^*_{xi}, \forall i$[33]. 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.

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

Nested logit demand

I modify the demand equation (6.2) in section 6.1 such that taxi and Uber are within one group allowing flexible substitution pattern.

$$\begin{aligned}\log(U^{ij}_{cft,pre}) &= \log(\tau^i_{ft}) + U^{ij}_{cft,post} \\ &= \underbrace{\log(\tau^i_{ft}) + \theta \ln(u^i_{xt}/v^i_{xt})\mathbb{1}_{f=x} + d_x + d_i + t + \xi^i_{ft}}_{\delta^i_{ft}} + \alpha^{ij} \ln(p^{ij}_{ft}) + \varepsilon^{ij}_{cft}\end{aligned} \tag{10.1}$$

Instead of assuming ε^{ij}_{cft} follows i.i.d. type I extreme value distribution, I assume that:

$$\varepsilon^{ij}_{cft} = \zeta^{ij}_{cgt} + (1-\beta)\nu^{ij}_{cft} \tag{10.2}$$

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 conditional market share on route ij at time t becomes product of share within group and share cross group. Within the group of taxi and Uber, the market share of taxi, for example, is:

$$s^{ij}_{y|gt} = \frac{\exp((\delta^i_{yt} + \alpha^{ij}\ln(p^{ij}_{yt}))/(1-\beta))}{\exp((\delta^i_{yt} + \alpha^{ij}\ln(p^{ij}_{yt}))/(1-\beta)) + \exp((\delta^i_{xt} + \alpha^{ij}\ln(p^{ij}_{xt}))/(1-\beta))} \tag{10.3}$$

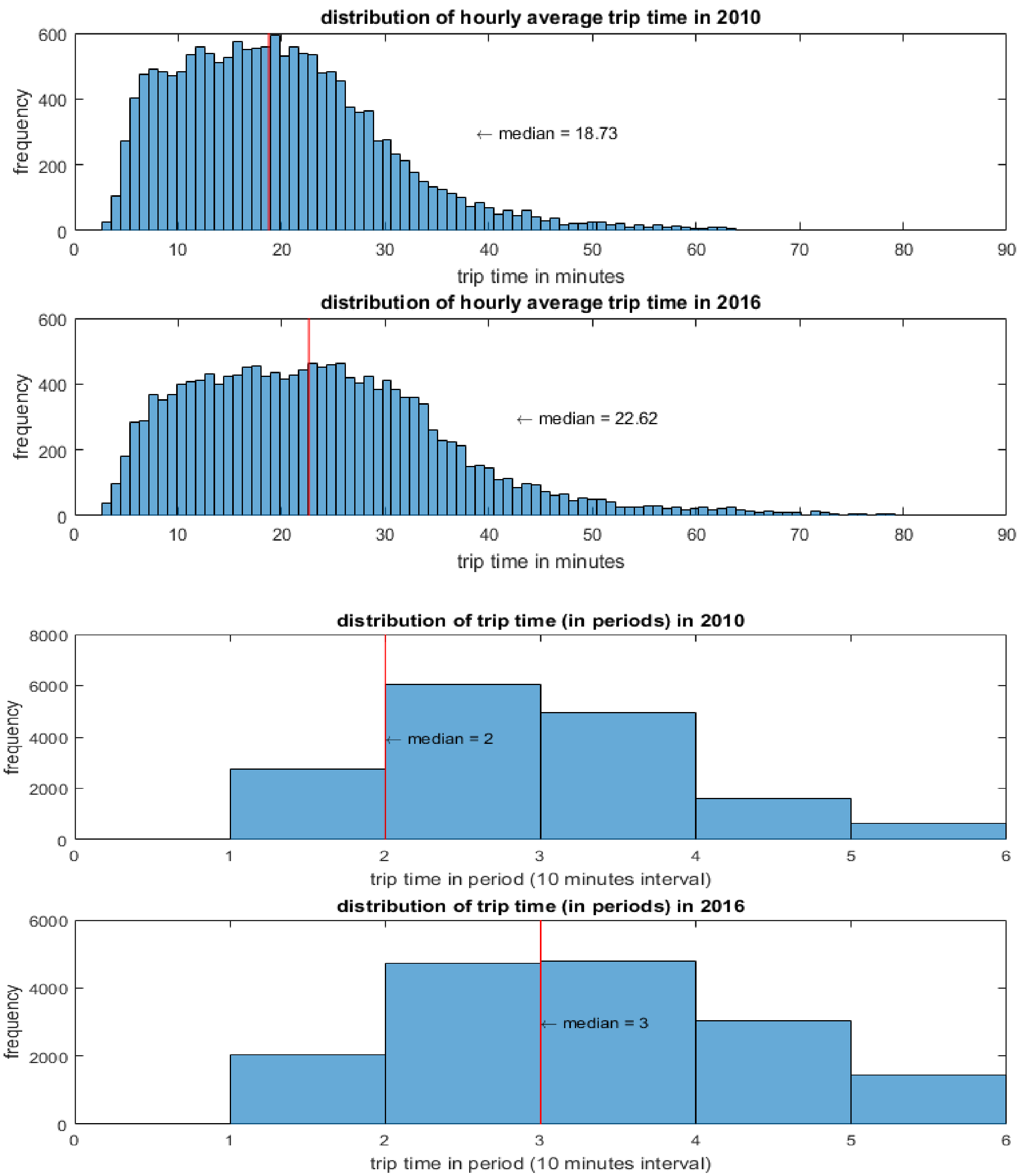
Denote the denominator as:

$$D_g = \exp((\delta^i_{yt} + \alpha^{ij}\ln(p^{ij}_{yt}))/(1-\beta)) + \exp((\delta^i_{xt} + \alpha^{ij}\ln(p^{ij}_{xt}))/(1-\beta)) \tag{10.4}$$

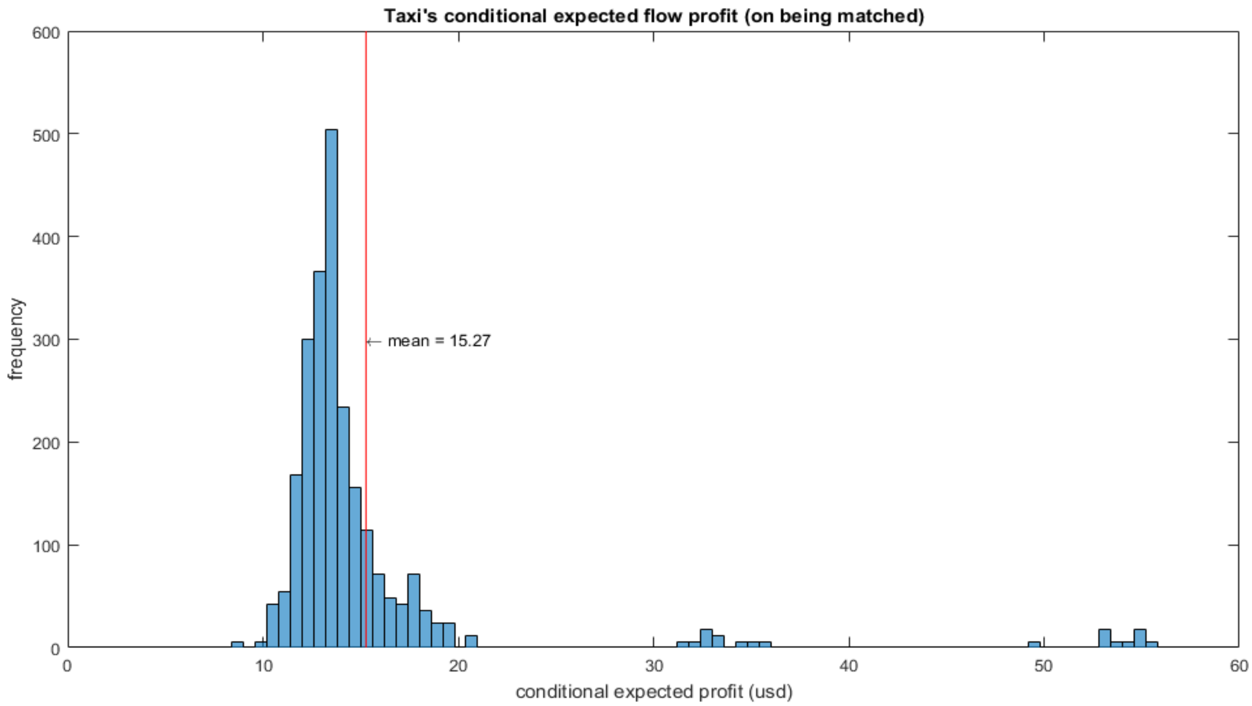
The probability of choosing the group of taxi and Uber is:

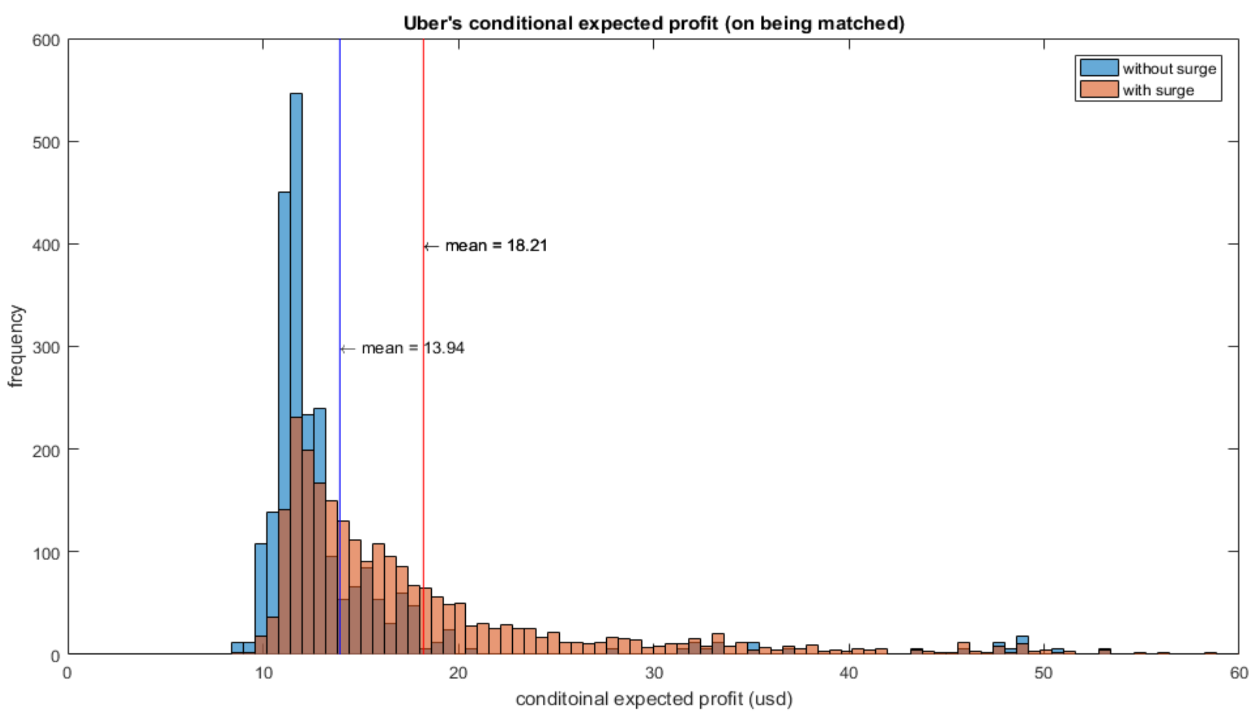
$$s^{ij}_{gt} = \frac{D_g^{1-\beta}}{1 + D_g^{1-\beta}} \tag{10.5}$$

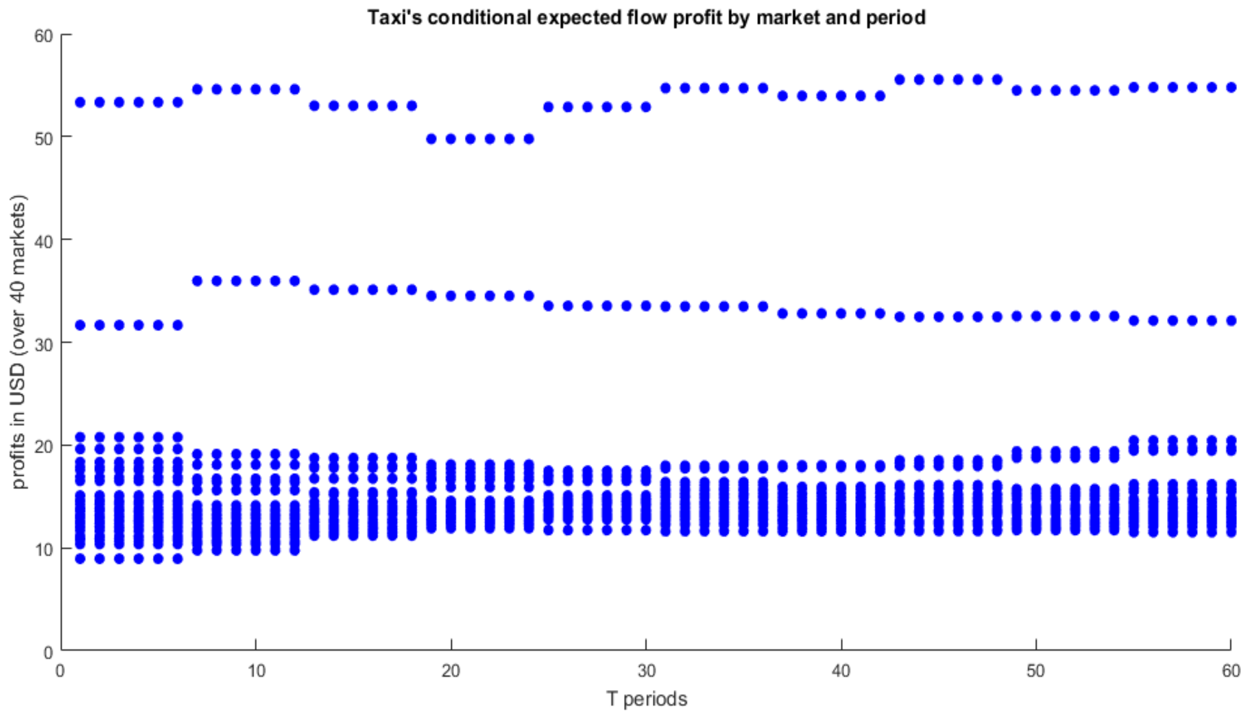
Then the conditional market share of (6.3) becomes $s^{ij}_{yt} = s^{ij}_{y|gt} * s^{ij}_{gt}$



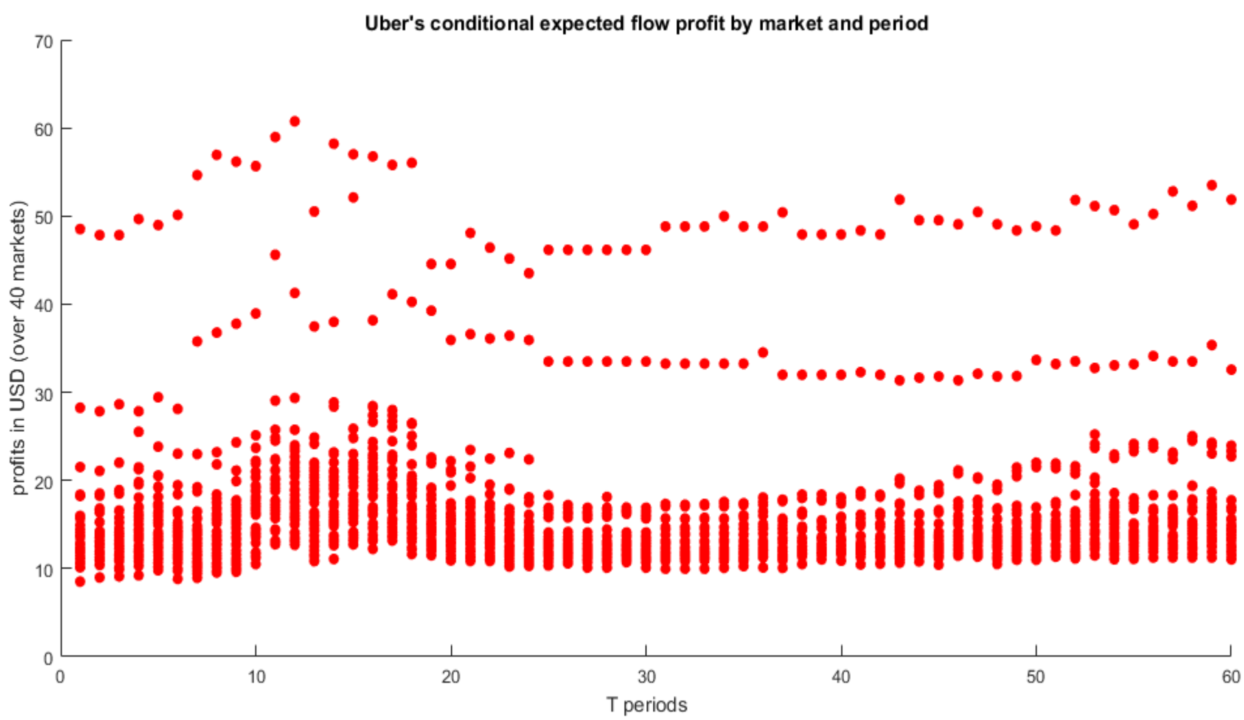
distribution of hourly average trip time in 2010
frequency
trip time in minutes
← median = 18.73
distribution of hourly average trip time in 2016
frequency
trip time in minutes
← median = 22.62
distribution of trip time (in periods) in 2010
frequency
trip time in period (10 minutes interval)
← median = 2
distribution of trip time (in periods) in 2016
frequency
trip time in period (10 minutes interval)
← median = 3


Taxi's conditional expected flow profit (on being matched)
600
500
400
300
200
100
0
frequency
mean = 15.27
0
10
20
30
40
50
60
conditional expected profit (usd)

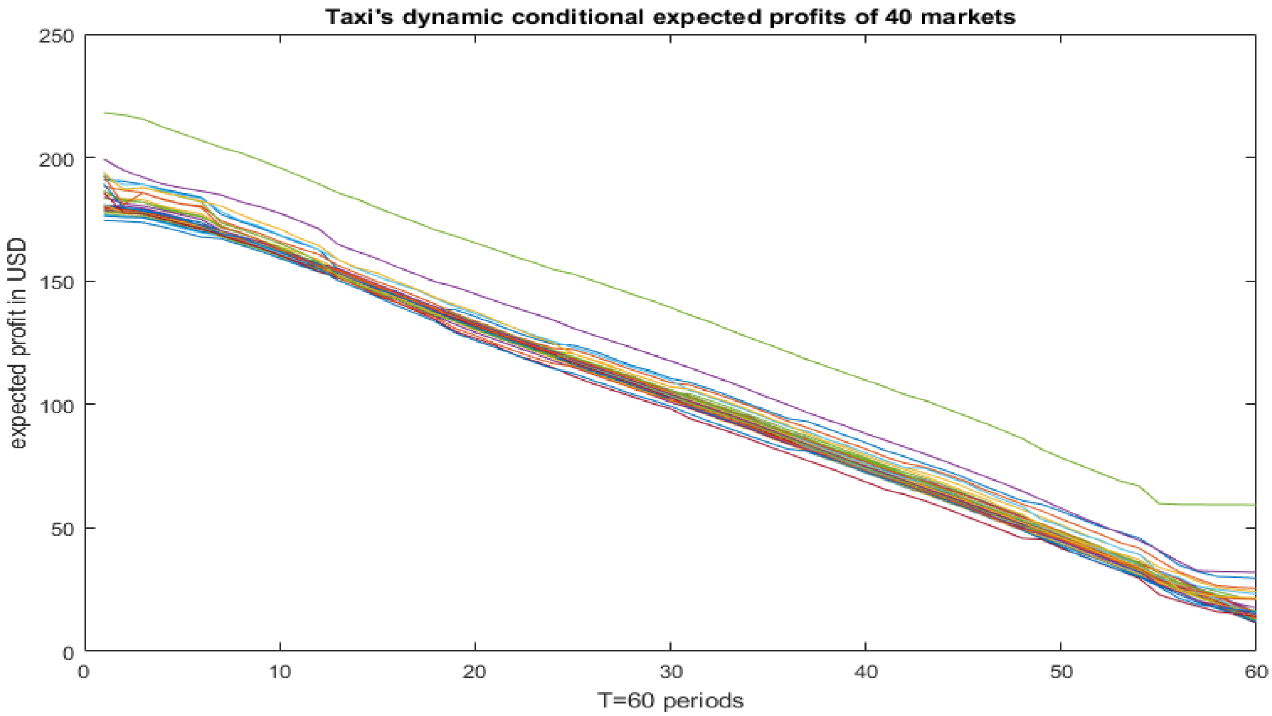

Uber's conditional expected profit (on being matched)
without surge
with surge
600
500
400
300
200
100
0
frequency
mean = 18.21
mean = 13.94
0
10
20
30
40
50
60
conditoinal expected profit (usd)



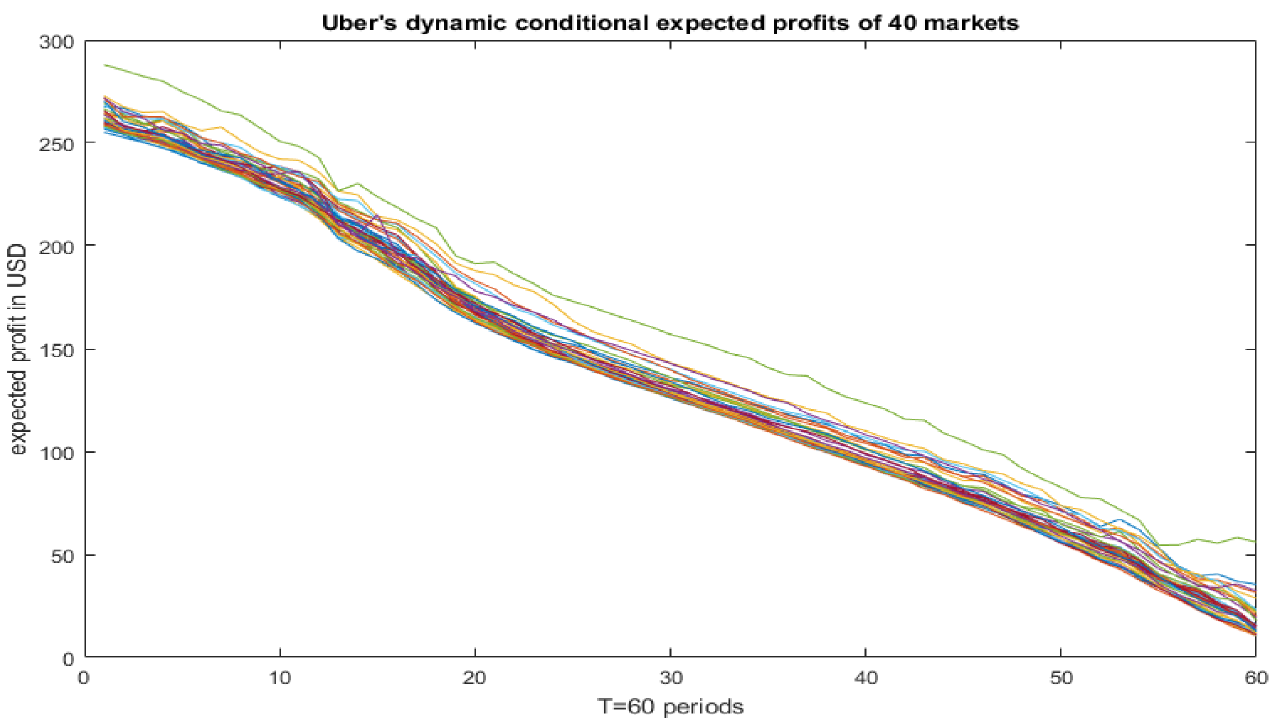
Taxi's conditional expected flow profit by market and period
profits in USD (over 40 markets)
T periods
0
10
20
30
40
50
60



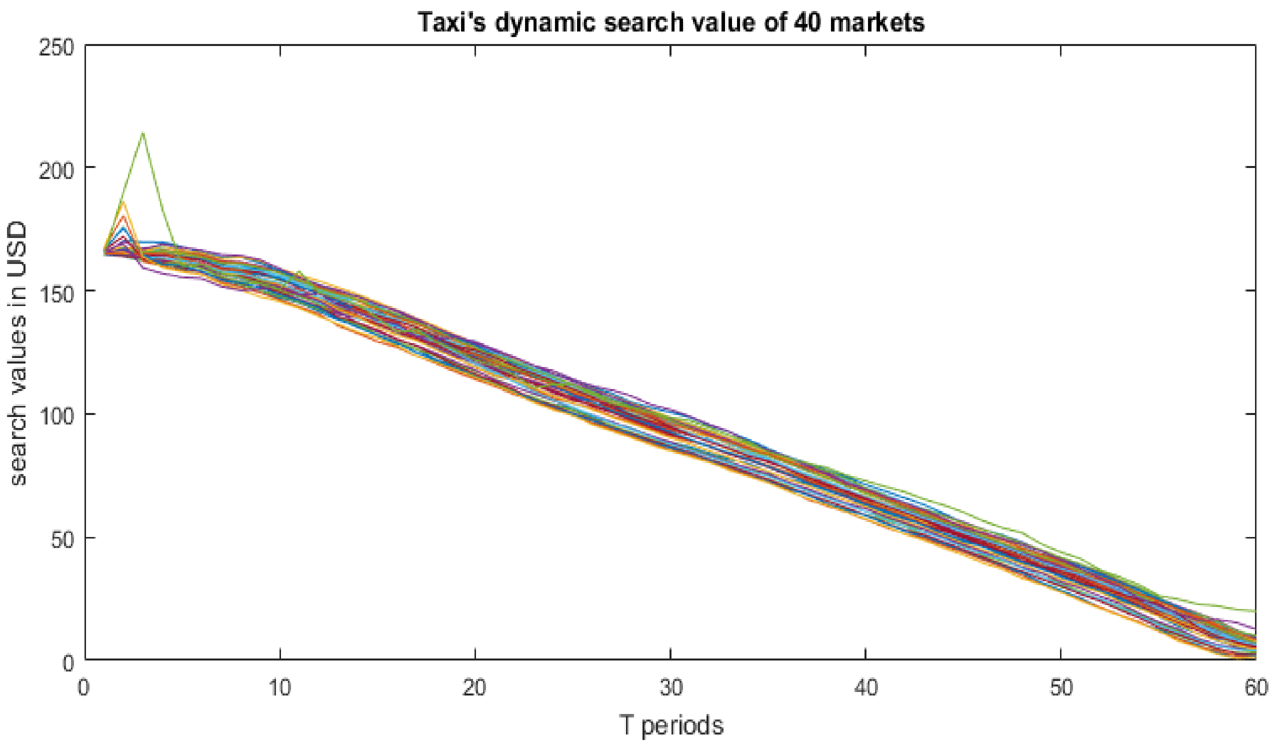
Uber's conditional expected flow profit by market and period
profits in USD (over 40 markets)
T periods
0
10
20
30
40
50
60
70

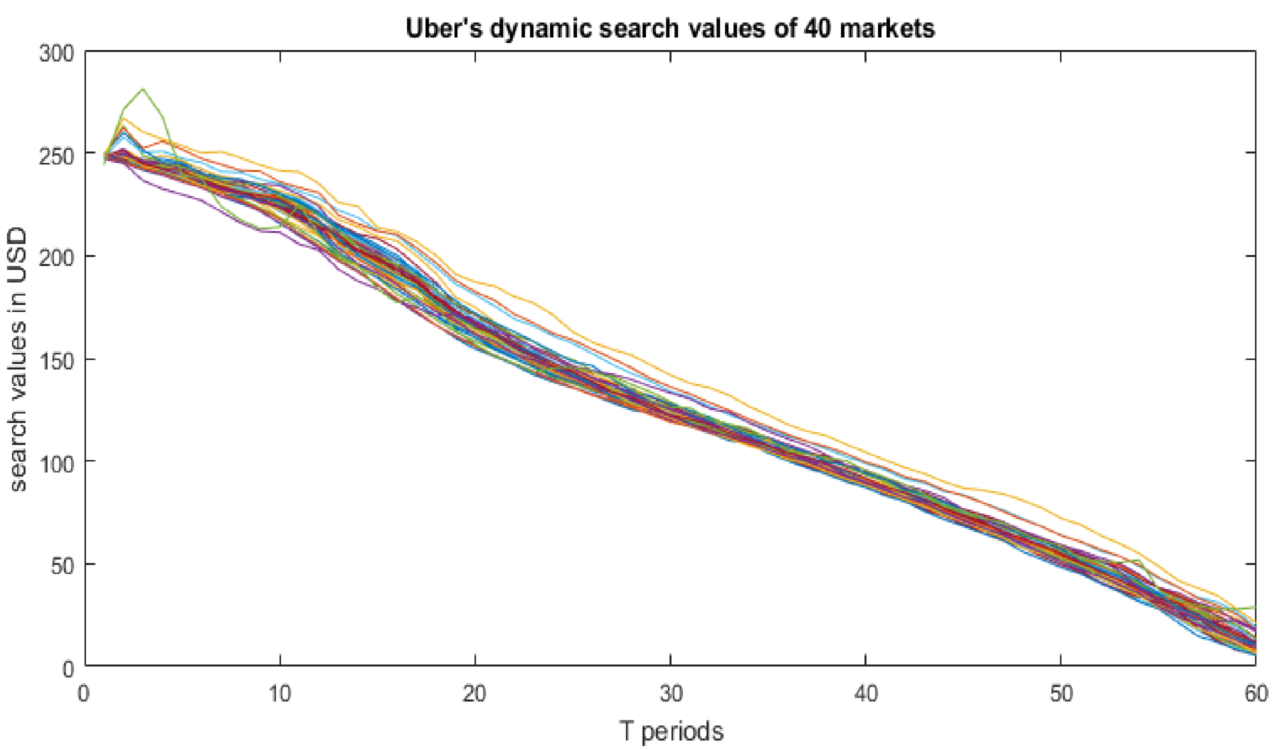


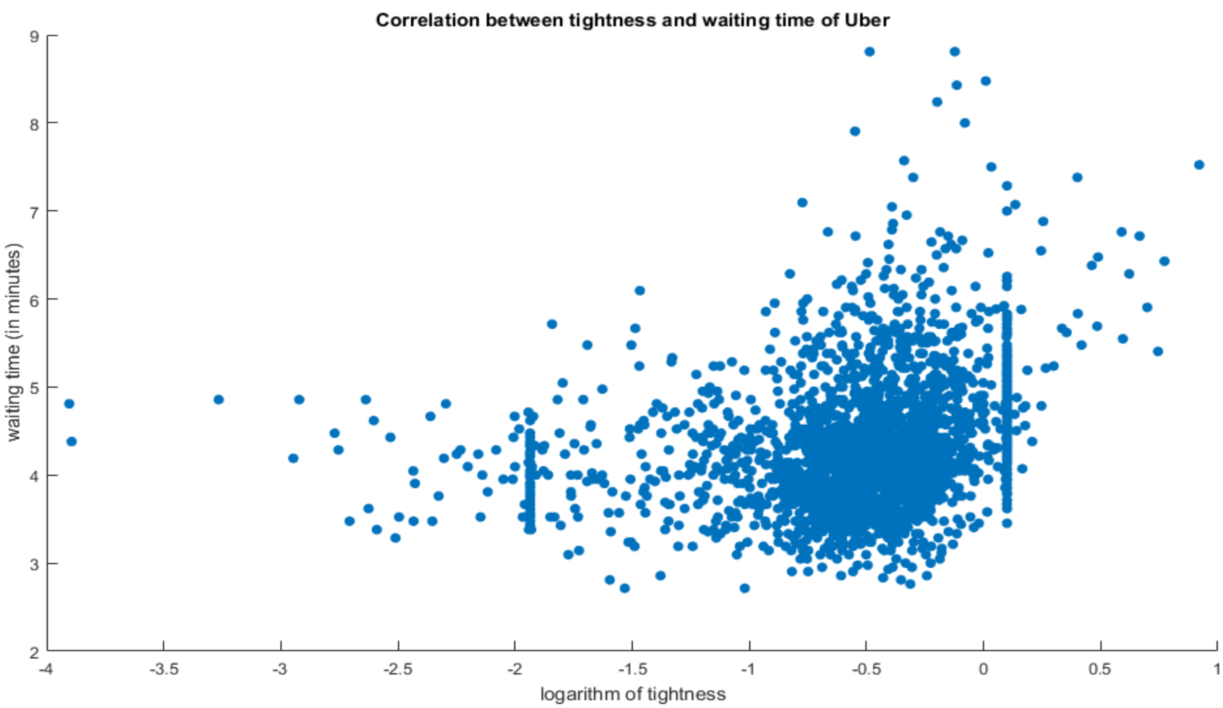
Taxi's dynamic conditional expected profits of 40 markets
expected profit in USD
T=60 periods
0
50
100
150
200
250
0
10
20
30
40
50
60



Uber's dynamic conditional expected profits of 40 markets
expected profit in USD
T=60 periods
0
50
100
150
200
250
300
0
10
20
30
40
50
60


Taxi's dynamic search value of 40 markets
search values in USD
T periods
0
50
100
150
200
250
0
10
20
30
40
50
60


Uber's dynamic search values of 40 markets
search values in USD
T periods
0
50
100
150
200
250
300
0
10
20
30
40
50
60


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