Competing Platforms and Transport Equilibrium: Evidence from New York City

Nicola Rosaia*

(Job Market Paper)

This version: November 19, 2020

Latest version available here

Abstract

This paper studies platform competition in the app-based transportation industry. I present a model of competing platforms in transport equilibrium, characterizing analytically the profit-maximizing allocations and prices, and estimate it using high-frequency data on the operations of the two main platforms in New York City. One main issue associated with the growth of these platforms is their impact on traffic volumes. I argue that this is exacerbated by missed economies of density from platform competition. I use the model to simulate the impact of a merger, finding that it would reduce the average number of idle vehicles in the Central Business District by 30%, improving efficiency and reducing vehicle traffic by 12%. I also use the model to study common policies aimed at reducing traffic volumes, such as congestion pricing and entry restrictions.

*Harvard University, Department of Economics, Littauer Center, Cambridge MA 02138, USA, nicolarosaia@g.harvard.edu
I am thankful to my advisors Myrto Kalouptsidi, Robin Lee, Ariel Pakes for constant advice and support, to Giulia Brancaccio for many helpful advices, and to Edward Glaeser, Gabriel Kreindler, Frank Pinter and Allen Zhang for helpful comments.

1 Introduction

Digital platforms connecting customers and independent service providers have become more and more prevalent in the economic landscape, one main example being the app-based transportation industry. Transportation Network Companies (TNCs) such as Uber and Lyft provide a large and growing share of vehicle trips in major cities in the United States, recently raising concerns about their impact on traffic in major urban areas.[1] Regulating digital platforms involves trade-offs due to the presence of "network effects", also known as "economies of density" in spatial contexts: on the one hand, users benefit from joining larger (or denser) platforms, taking advantage from a larger (or denser) pool of trading partners; on the other hand, concentration comes at the cost of higher platform market power. Moreover, to be competitive, platforms must achieve a large enough scale so to realize economies of density at a sufficient level. In the context of transportation, this generates complex trade-offs between competition, economic efficiency, and the externalities imposed by traffic on the rest of society. This paper seeks to measure these trade-offs in New York City, one of the world's largest markets for TNCs, generating more than 15 million trips each month, dispatching more than 80,000 drivers.

To do so, I collect high-frequency data on prices, waiting times, and drivers' dynamic geolocations throughout a dense grid of observation points, and match them with publicly accessible trip records, obtaining a complete picture of the operations of the two main platforms during a period of two months.

Building on this data, I develop a model of competing platforms in transport markets. The model has two main layers: a dynamic passenger-driver economy, and an oligopolistic sector of profit-maximizing platforms competing to attract both types of users. The first layer is described by means of a transport equilibrium model: on the demand side, passengers seek transportation between different regions of the city, facing waiting times which depend on the nearby density of drivers available on different platforms; on the supply side, drivers solve a dynamic problem where they enter, exit, switch platforms, move between regions searching for trips requests, and transport passengers across the city; the spatial allocation of demand and supply across platforms is determined by a dynamic equilibrium of this process. Spatial models of this kind, dating back to Lagos [2000], have recently become popular in the transportation literature, allowing to capture central features such as geographic spill-overs and the spatial frictions inherent to the

[1] Congestion taxes on ride-hail are currently in place in Chicago, New York City, Seattle and under discussion in San Francisco.

matching process, while prices are usually either taken as given, or determined by means of decentralized bargaining.[2] To model the market's endogenous price response to changes in the environment, I nest on top of this structure an oligopolistic sector of profit-maximizing platforms, implementing equilibrium allocations by setting dynamic prices and compensations for trips between different regions. Platforms seek to attract users on both sides as in Rochet and Tirole [2003], and can avoid coordination failures, but complexities arise from the dynamics and the geographic spill-overs.[3]

I derive an implementation cost function pinning down drivers' equilibrium compensations at any given allocation, and use it to characterize platforms' best response as an allocation choice problem. Intuitively, platforms choose the quantity and "quality" of trips on different routes, where the latter is captured by the car supply in different regions, which in turn determines passengers' value for the service through their expected waiting times. Platforms distort à la Spence [1975], under-providing quantities, and internalizing only the value of quality for marginal customers (the Spence distortion).[4]

This characterization yields simple computational algorithms, and an analytical expression for the profit-maximizing two-sided prices. Drivers' compensations are set so to align their expected dynamic matching surpluses with platforms' marginal revenues from quality in different regions, while passenger prices are the sum of three components: a "static" marginal cost based on time and distance, a classic market power distortion, and a dynamic component trading off drivers' marginal profitability at origin with the marginal benefit of expanding supply at destination.

I structurally estimate the model combining variation from different sources.[5] On the supply side I estimate drivers' dynamic discrete choice model exploiting variation in the number of available drivers in time and space. Standard methodologies for estimating dynamic models rely on observing the frequencies at which agents make alternative choices in different states (i.e. agents' conditional choice probabilities).[6]

[2]Spatial models have been used, for instance, to study search frictions in taxi markets (Buchholz [2020] and Frechette et al. [2019]) and oceanic transportation (Brancaccio et al. [2020b]), the role of network effects on trade costs (Brancaccio et al. [2020a]), welfare improvements from centralization (Shapiro [2018] and Liu et al. [2019]), demand and supply imbalances (Ghili and Kumar [2020]), and the welfare effects of dynamic pricing Castillo [2019].

[3]Coordination failures in two-sided markets might arise when users do not join the platform because they expect users on the other side to not join as well (and these expectations are self-fulfilled), but these are usually assumed away. of See for instance Armstrong [2006] and Weyl [2010].

[4]The fact that two-sided platforms might generate this type of distortions in a static pricing model has been previously noted by Weyl [2010].

[5]Papers estimating demand and supply elasticities on specific ride-sharing platforms include Cohen et al. [2016a], Buchholz et al. [2020], Angrist et al. [2017], and Castillo [2019], among others. Other papers have studied specific aspects of labor supply (e.g. Hall et al. [2019], Chen et al. [2019], Cook et al. [2018]).

[6]See for instance Pakes [1986] and Rust [1987] for single agent models and Bajari et al. [2007] and Pakes et al. [2007] for

This requirement is not easily met in studies of this type of markets, where usually drivers' actions (i.e. decisions to enter/exit in different regions and relocation movements) are not observed. On the other hand, my data allows to measure the frequency at which drivers *visit different states*, i.e., the frequency at which they are available on different platforms in different regions. I show that these frequencies can be inverted for drivers' equilibrium payoffs, provide a simple algorithm implementing this inversion by means of successive tâtonnements, and exploit it to estimate drivers' parameters.

On the demand side, I estimate a nested logit model of discrete choice of transport mode, relying on high-frequency variation in prices, waiting times and trip requests, and dealing with endogeneity by exploiting a policy change as a natural experiment.

Finally, I recover a numerical representation of platforms' matching technology by simulating a function associating different densities of idle drivers to passenger waiting times in different regions, and validate it from the observation of platform-reported waiting times.[7]

I use the estimated model to simulate the market equilibria under different scenarios. In the first counterfactual, I simulate the impact of a merger. This allows me to quantify the competitive trade-offs due the the presence of economies of density: customers' value for the service depends on waiting times determined by the nearby density of idle drivers, while drivers' wages increase with the density of customers. This raises the question of whether more competition is efficient, as it comes at the cost of lower density within each platform. This question is of broader relevance than the ride-sharing industry, speaking to the literature on network externalities and two-sided markets, and to the emerging literature on competition between digital platforms.[8] I find that merging the two platforms would produce substantial efficiency gains, of about 150 M$ per year, although all gains are captured by the monopolist through increased profits, leaving users on both sides worse off. This crucially relies on several factors, such as passengers' sensitivity to prices and waiting times, as well as the level of concentration and market density and prior to the merger, as it can bee seen by looking at the variation of the effects across regions. In particular, the negative effect on customer welfare is inversely related to the previous market density.

strategic setups.

[7] Frechette et al. [2019] use a similar idea to simulate a matching function for NYC taxicabs.

[8] Several papers study competition in traditional two-sided markets, such as yellow pages (Rysman [2004]), software and video games (e.g. Clements and Ohashi [2005], Lee [2013], Corts and Lederman [2009], Gandal et al. [2000]), newspapers and radio industry (e.g. Chandra and Collard-Wexler [2009], Kaiser and Wright [2006], Jeziorski [2014]), among others, while Cao et al. [2018] study platform competition in bike-sharing.

This is because the monopolist can profitably expand car supply in less dense regions, where the benefit for customers of reduced waiting times more than offsets the price increase.

This inefficiency has significant effects on traffic volumes. I find that merging the two platforms would reduce the average number of TNC vehicles on the streets by 12%. This is due to the fact that, in order to be profitable, both TNCs must achieve a large enough size so to achieve sufficient density. On the other hand, the monopolist makes a more efficient use of drivers, as reflected by the fact that the average number of idle drivers in the city drops by 26%, while passenger trips decrease by only 1.4%.

In the second and third counterfactuals, I compare congestion pricing schemes targeting fixed and variable costs of accessing the Central Business District, by simulating the effects of charging drivers on a per-minute basis, and comparing these with the effects of charging drivers a fixed amount at every entry. I find that fees targeting fixed costs, such as the NYC's proposed congestion pricing plan, have small effects on TNC traffic.[9] Intuitively, this is a consequence of the fact that the CBD is a "net attractor" of trips during daytime, meaning that it receives more trips than those it generates. Hence car supply in the CBD draws mainly on drivers who remain idle after drop-off, rather than on a net inflow of idle drivers from outer regions. In contrast, per-minute fees can have a significant effect, by providing platforms with the right incentives to reduce vehicles' idling time. In particular, charging drivers 5$ per hour driven in the CBD would reduce traffic by 8.5%, and idling time by 20%, with mild effects on customer surplus.

In the last counterfactual I study the impact of entry restrictions, by simulating the impact of increasing drivers' licensing costs to 10,000$ per year. As a result, the number of licensed drivers drops to about 13,500, providing a basis for comparison with the medallion system (the total number of medallions is currently 13,587 in NYC). Economies of density provide an argument against the introduction of this type of policies.[10] In particular, it is often argued that these would have the drawback of excluding from service peripheral regions where economies of density are harder to realize, while inducing platforms to concentrate activities in the denser, most congested regions. Interestingly, I find that this is not the case, but instead platforms are able to coordinate the smaller pool of drivers so to avoid a concentration of activities in central regions. I use the results of the simulation to argue that economies of density are not crucial to understand the geographical variation in the impact of such policies, which can instead be

[9] In a recent paper, Kreindler [2020] studies Peak-Hour Road Congestion Pricing. It is important to stress that my results concern the traffic generated by TNCs, and not by private vehicles.

[10] In a recent paper, Frechette et al. [2019] made this case for entry restrictions on NYC taxicabs.

explained by looking at the network of travel patterns. In the aggregate, the policy generates a sizable welfare loss, of about 200 M$ per year, reducing TNC traffic by about 7%.

Outline

The paper is structured as follows. Section 2 describes the industry and the data used; Section 3 provides some some descriptive evidence; Section 4 presents the model; Section 5 describes the estimation; Section 6 presents the counterfactuals; Section 7 concludes.

2 Market and Data

2.1 Market Overview

The app-based segment of the for-hire vehicle industry (ride-sharing) in New York City is operated by four companies, whose services consist in connecting independent drivers with passengers through smartphones.[11] Uber introduced its flagship service UberX in 2012, subsequently joined by Via in 2013, Lyft in 2014, and Juno in 2016. The market has expanded rapidly since 2012, including about 85,000 drivers, dispatching more than 15 million trips each month. This rapid growth has generated benefits for riders by means of low fares and the extension of taxi services to peripheral regions. It has also created new jobs for drivers, whose majority today is made of full-time workers, independent contractors setting their own schedule and providing their own vehicles.[12] On the other hand, it has diverted passengers from mass transit, contributing to substantial declines in New York City subway and bus ridership.[13] It has also contributed to the worsening of congestion in the Manhattan Central Business District (CBD), summing with other factors such as tourism and population growth.[14] To tackle these issues, the city implemented a congestion fee on ride-sharing trips, and authorized a congestion pricing plan to be implemented in 2022. As a temporary measure, the city also limited the extension of new for-hire-vehicle licenses.

[11] As of today, Juno has been acquired by Lyft.

[12] See Parrott and Reich [2018].

[13] In its 2017, 2018 and 2019 Citywide Mobility Surveys of over 3,300 residents each, the NYC Department of Transportation found that app trips are more often replacing transit than any other mode of travel. For cities in the United States, Graehler et al. [2019] found that in the TNCs entry is usually associated with a significant decrease in transit ridership.

[14] In a recent paper, Molnar and Mangrum [2018] analyze of the impact of TNCs on congestion in NYC in depth.

2.1.1 Policy change

In my empirical strategy, I take advantage of two policies that went into effect in New York City on February 1, 2019: a 2.75$ tax on all ride-sharing trips crossing the CBD, and a minimum pay standard for ride-sharing drivers. The latter established a minimum payment per mile and per minute spent transporting passengers, which is adjusted quarterly in order to ensure that drivers receive an overall wage of at least 17.22$ per hour.

2.1.2 Proposed Congestion Pricing Plan

The framework for congestion pricing was implemented in the 2019 New York State budget. It consists in a fee, whose exact amount is still to be decided, that will charge all vehicles traveling into the CBD. This will go into effect by 2022, and will be the first congestion pricing scheme enacted in the United States.

2.2 Data

I collect data on the two largest platforms operating in the market, covering more than 90% of the total market share, measured in passenger trips. In what follows, I will refer to them as platform a and platform b. I gather information on the operations of these platforms by leveraging their client apps.[15] After a user opens the app and authenticates with the platform, the app sends messages to the platform's server every few seconds. Each message includes the geographic coordinates of the user's origin and desired destination, and the server responds with a JSON-encoded list of information which includes: a price estimate; the expected waiting time; a list of the nearest eight drivers available to receive trip requests; information about the surge multiplier at origin, time and distance of the trip, which can be used to calculate drivers' compensations.[16] Each driver is represented by a unique identifier and a list of GPS coordinates and timestamps tracing its recent trajectory.[17] I write scripts emulating the exact behavior of the apps, sending messages every few seconds and recording the responses. By controlling the coordinates

[15]This type of data has already been used in previous studies, informing official reports by the San Francisco County Transportation Authority. See Chen et al. [2015], Jiang et al. [2018], SFCTA [2017] and Castiglione et al. [2018].

[16]During my sample period, drivers' compensations were obtained as the product of a component based on time and distance and the surge multiplier. The latter is a dynamic origin-specific factor that "clears the market", changing depending on demand and supply conditions.

[17]These identifiers are randomized each time a car becomes available, hence they do not allow me to track individual drivers over time.

sent by the script, I can collect data for arbitrary origin-destination pairs.

2.2.1 High-Frequency Data

I cover a vast geographic area with observation points. To make sure that I record the trajectories of all idle drivers, I let the space between two points be less than the minimum distance to the furthest car they observed over a two-week experimentation period. Each point makes a request every 20 seconds for two months, from May 15, 2019 to July 16, 2019, recording the trajectories of the eight closest cars, waiting times, surge multipliers, and the price quotes to different destinations, where the destination rotates among a fixed grid of locations. I match this data with the records of origin, destination, time of pick-up and time of drop-off all ride-sharing trips in the the same period, which are published monthly by the Taxi and Limousine Commission (TLC).

To construct dynamic estimates of different variables, I discretize time in 10-minute intervals, and space by dividing the area in 39 regions obtained by aggregating the taxi-zones defined by the TLC. For each interval, I compute the average passenger price, waiting time, driver compensation, time and distance for trips between every pair of regions. I also measure car supply in each region, by simply counting the total number of unique cars observed across all observation points. Finally, I exploit drivers' trajectories to construct dynamic estimates of the traffic speed in different regions.

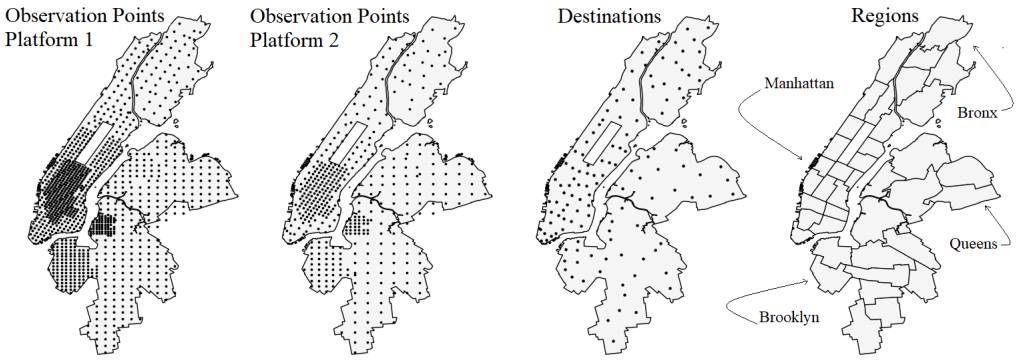

Figure 1: Observation points represent the locations from which data requests are made. The distance between them reflects the observed density of idle drivers. Hence points are closer in central regions than in the periphery, and closer for the larger platform. Each point requests price quotes rotating among different destinations. Regions represent my geographical subdivision of the market, obtained by aggregating the TLC's taxi zones.

2.2.2 Pre-post Policy Data

I also collect data at lower frequency before and after the policy change described in Section 2.1.1, from January 25, 2019 until February 8, 2019. During this period, I collect the waiting times and price quotes for trips between every pair of destinations in Figure 1 every 10 minutes. I then average these observations to construct hourly estimates of the average waiting time and passenger price for trips between every pair of regions in my geographical subdivision. Matching this data with the TLC trip records for the same period allows me to measure changes in prices, waiting times and trip requests at the weekday, hour and origin-destination level for the weeks before and after the policy.

2.2.3 MTA Travel Survey

I complement this dataset with information on the travel behavior of New Yorkers from the 2008 New York Customer Travel Survey conducted by Metropolitan Transportation Authority (MTA).[18] This survey interviewed a random sample 16,186 NYC residents May through November 2008. It collected data on time, origin, destination, and other characteristics of all trips each member of the respondent's household took during the day previous to the interview. Weighting for economic and demographic characteristics of the households, I expand this data to obtain estimates, at the weekday and hour level, of the total number of trips undertook by NYC residents on all transport modes, between different regions in my geographical subdivision. In this way, I can obtain estimates of the share of ride-sharing trips out of the total number of trips on all modes.

2.2.4 Other Information

I integrate this information with the TLC's Monthly Data Reports, from which I obtain estimates of the total number of unique ride-sharing drivers (the total number of unique drivers who provided at least one trip in the course of a month). Finally, I use routing information provided by the client apps to obtain estimates of the average time and distance of trips between all regions in my geographical subdivision, at the weekday-hour level.

[18] As soon as it becomes available, I will substitute this with data from an analogous survey conducted in 2019.

3 Descriptive Evidence

In this section I provide some descriptive evidence about the functioning of the market; my findings motivate the model and counterfactuals. I start by presenting some statistics of the intra-day variation in prices, wages, waiting times and profit margins. I then document the intra-day and geographical variation in car supply, and discuss the mechanism generating this variation. I conclude by describing the matching process and simulation, and by analyzing drivers' multi-homing behavior.

3.1 Prices, Wages and Match Qualities

A distinctive feature of this market is that platforms adjust passenger prices and driver compensations dynamically. This variation is partly due to platforms' objective to achieve market clearing, but is also explained by variation in platforms' profit margins. Figure 2 displays the patterns in prices, average mark-ups and driver hourly wages.

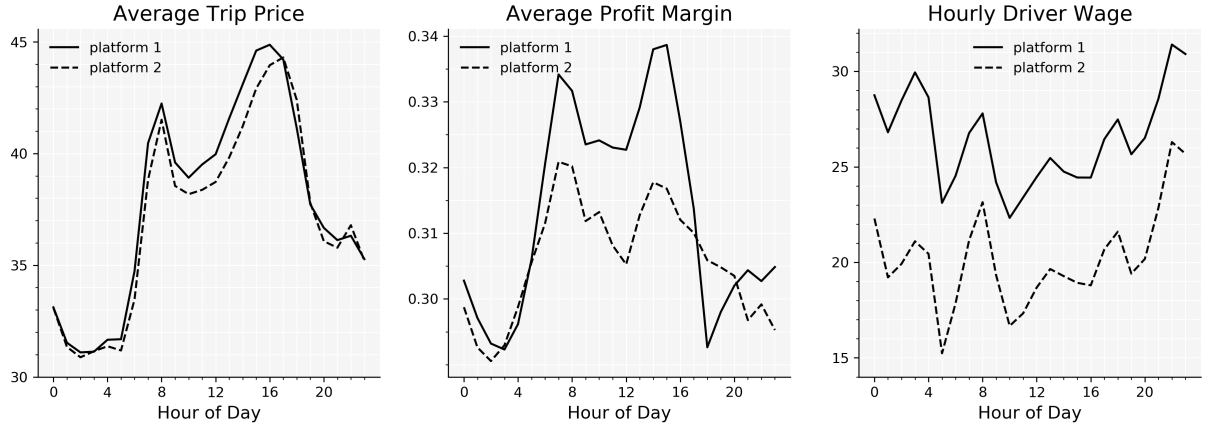

Figure 2: Intra-day patterns of prices, profit margins and hourly wages. Prices and profit margins are **unweighted** averages across all origin-destination pairs. Hence they reflect dynamic variation in platforms' pricing behavior and not intra-day changes in travel patterns. The average hourly wage is computed as total driver compensations over total drivers' working time.

Platforms charge an average mark-up of about 30%, while driver wages are usually around 23$ per hour. Figure 3 displays the intra-day variation in passenger waiting times and drivers' utilization rates, defined as the share of time during which drivers are matched with a passenger out of the total time they are active on different platforms

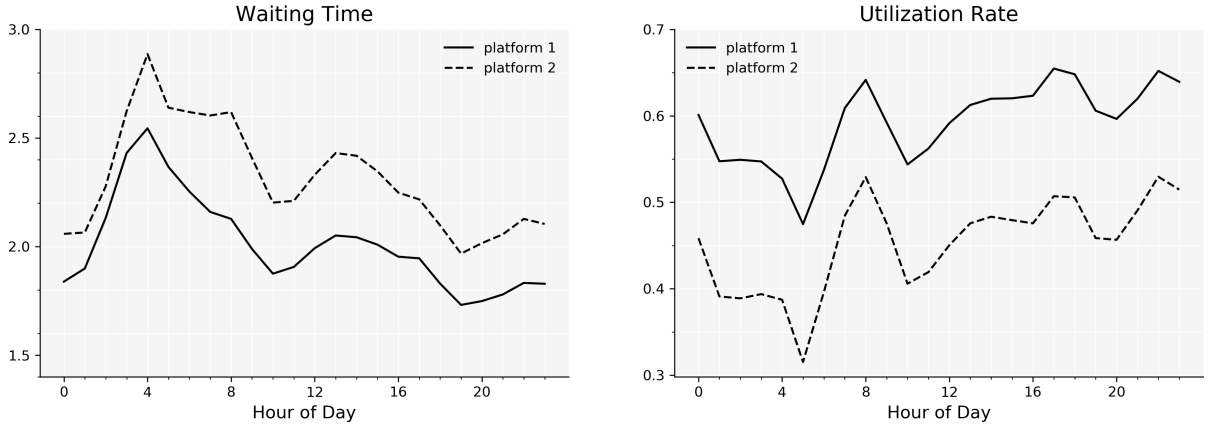

Figure 3: The left panel shows the average passenger waiting time, in minutes, averaged across trips. The right panel shows drivers' utilization rate, defined as the time that drivers spend matched with a passenger as a fraction of the total time they spend working.

Waiting times and utilization rates are quite stable during the day, indicating an effective coordination between demand and supply. Comparing with Figure 2, variations in driver wages are not explained by variations in the utilization rate, leaving room for variation in drivers' compensations. Interestingly, waiting times and utilization rates are inversely related, indicating the presence of economies of density. The average waiting time is very low, ranging between 2 and 3 minutes. On the other hand, driver utilization rates are also low, ranging between 30% and 65%.

3.2 Car Supply and its Determinants

Figure 4 displays the average supply patterns in a typical day (including weekends). In particular, it shows direct estimates of the number of idle drivers, obtained from GPS trajectories as the average total number of cars observed in a given instant. From these estimates and the number of trip requests, using information on waiting and travel times, I can also compute the total number of active drivers that are either idle or matched with a passenger at any given instant. The first thing to note is that drivers' activity is generally well below capacity (during my sample period there were about 85000 active ride-sharing drivers in New York City). Second, substantial intra-day variation in drivers' activity exists.

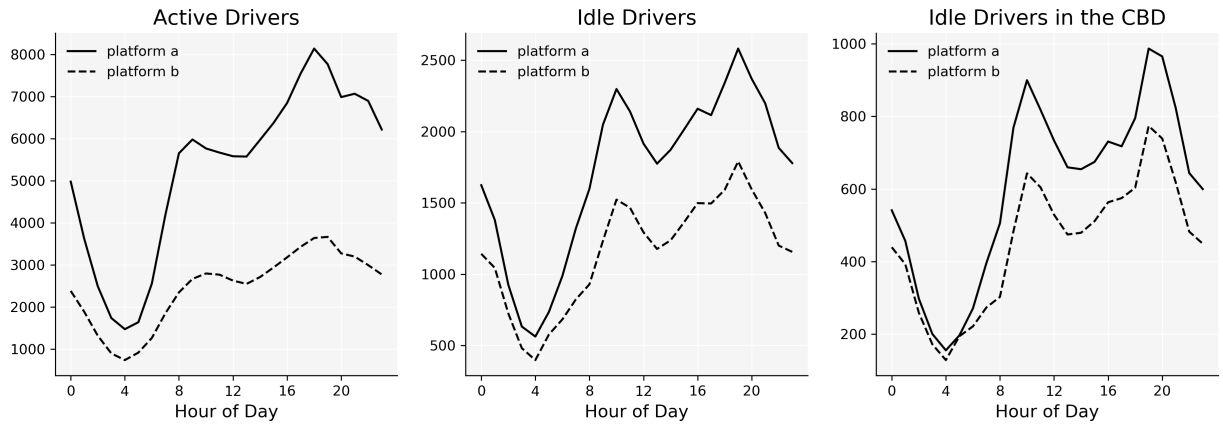

Figure 4: Intra-day supply patterns. The number of idle drivers is obtained as the average number of idle drivers observed in a typical instant from any observation point. Using information on trip requests and average travel and waiting times, I compute the average number of drivers that are matched with a passenger. Active drivers are obtained as the sum of idle and matched drivers.

The majority of vehicles are observed in the CBD, generating a substantial amount of traffic. Assuming an average speed of 7 miles per hour, TNC drivers account for about 508,000 miles driven in the CBD per day, 178,000 of which are due to empty vehicles.

There are two main determinants of supply in different regions: net arrivals and relocation movements of idle drivers. Net arrivals are defined as the number of drop-offs minus the number of pick-ups. Intuitively, a region receiving more trips than those it generates can draw from many drivers who remain idle after drop-off. The geographic distribution of net arrivals and its intra-day variation in the CBD are displayed in Figure 5. Notice that the CBD is a "net attractor" of trips during daytime. This implies that car supply in this region largely draws from drivers who remain idle after drop-off, and it does not need to be actively supported by drivers relocating inside. This observation has implications for the effects of congestion pricing, as I will argue in Section 6.2. Travel patterns also generate the need for drivers to actively relocate across regions in order to meet passenger needs.

Figure 6 displays the relationship between drivers' relocation patterns and net arrivals, and that the final balance between these two determines the net total inflow of drivers in different regions, hence the variation in car supply.

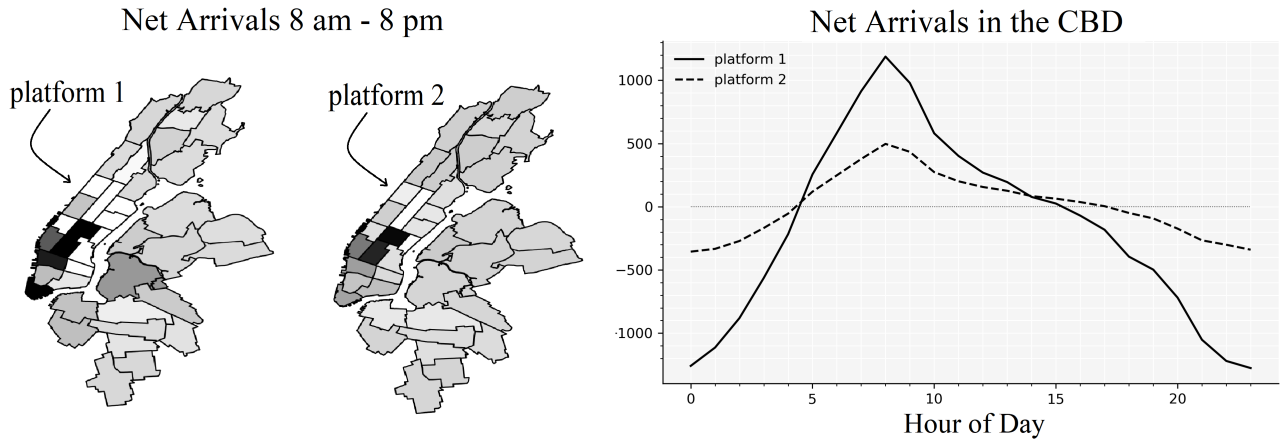

Figure 5: Net arrivals per hour, defined as the difference between the number of incoming trips and the number of outgoing trips. The right panel shown the average by hour of day in the Manhattan CBD. The maps show the distribution in different regions: darker regions are associated with more net arrivals.

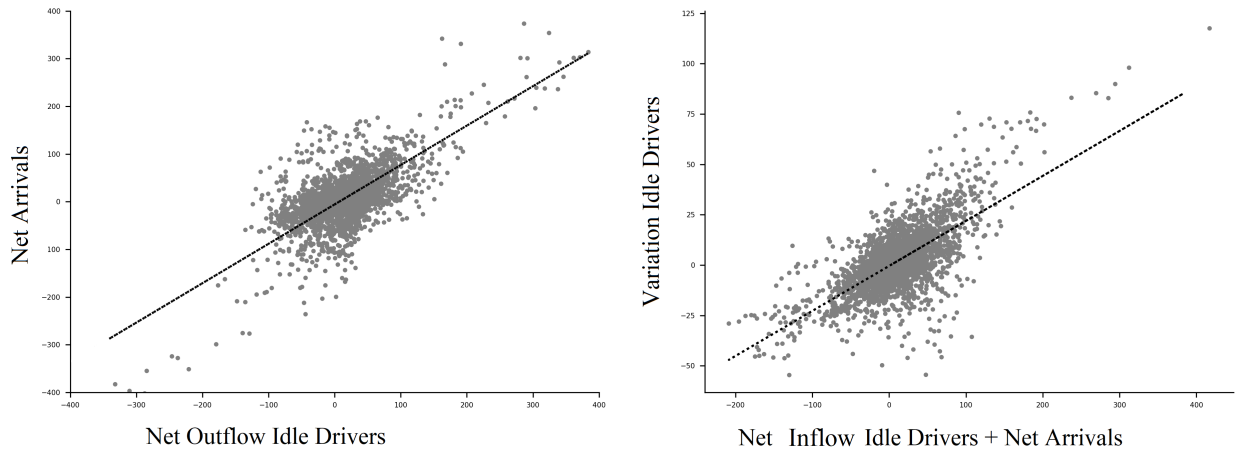

Figure 6 : In the left panel, the horizontal axis measures the net outflow of idle drivers in different regions, obtained from the GPS trajectories, averaged across hours of the day. The vertical axis measures the net arrivals of trips (difference between arrivals and departures), again averaged across hours of day. In the right panel, the horizontal axis measures the total net inflow of drivers in different regions, obtained as the sum of net inflow of idle drivers and net arrivals, averaged across hours of day. The vertical axis measures the average hourly changes in the number of idle drivers in different region.

3.3 Matching

In this section I explore the relationship between the waiting time that a customer faces, and the total number of available idle drivers nearby. This is important to quantify, since it affects platforms' incentives to maintain a more or less large pool of idle drivers in order to affect customers' valuations. I use my data

to obtain a non-parametric estimate of this relationship, which will be later be an input of the model.

When they receive a trip request, platforms immediately match it with the closest driver available.[19] It then takes some time for the driver to meet the passenger, depending on the travel time between them, and meanwhile the passenger must wait. For any given number of drivers, I simulate different configurations of their positions and the position of a passenger in the region as in Figure 7. For each configuration I compute the driving time between the passenger and the closest driver, depending on their distance and on the observed traffic speed. I then average the driving time across a large number of simulated configurations to find the expected waiting time.

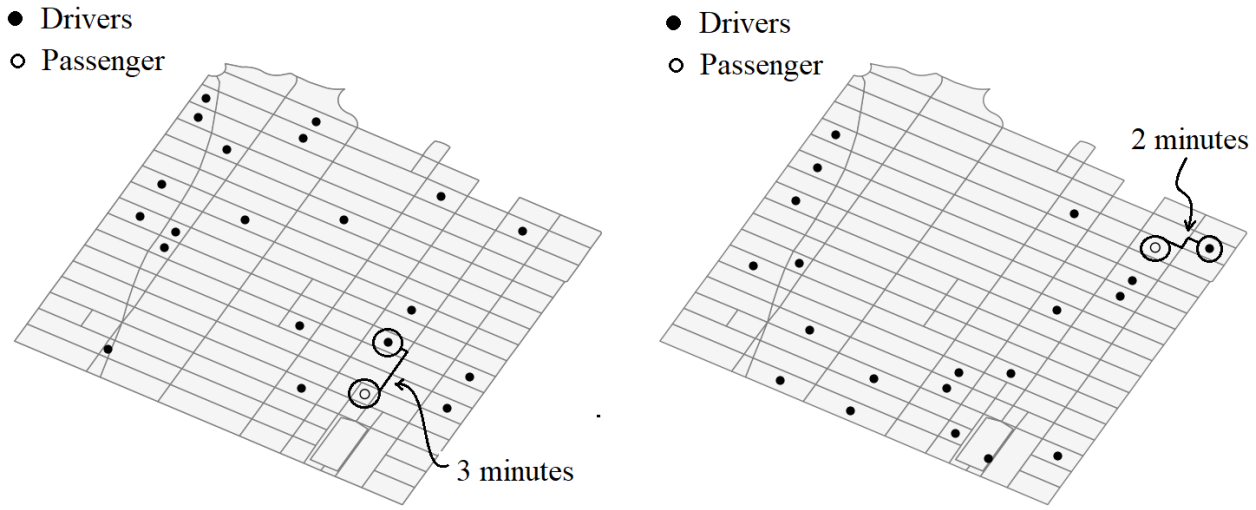

Figure 7: A graphical representation of the simulation of the matching function in a region. It displays two random configurations of 20 idle drivers and a passenger, and the associated travel times. The average waiting time corresponding to a given number number of idle drivers is obtained by averaging the travel times across a large number of random configurations.

In this way, fo every region, I construct a function associating an expected waiting time to any number of idle drivers. To validate this procedure, Figure 8 compares the simulated waiting times with the average waiting times reported by platforms in different regions at different hours and weekdays.

[19] Platforms later switched to more sophisticated technologies, where trip requests are let waiting for a short time interval and then matched in batches with nearby drivers.

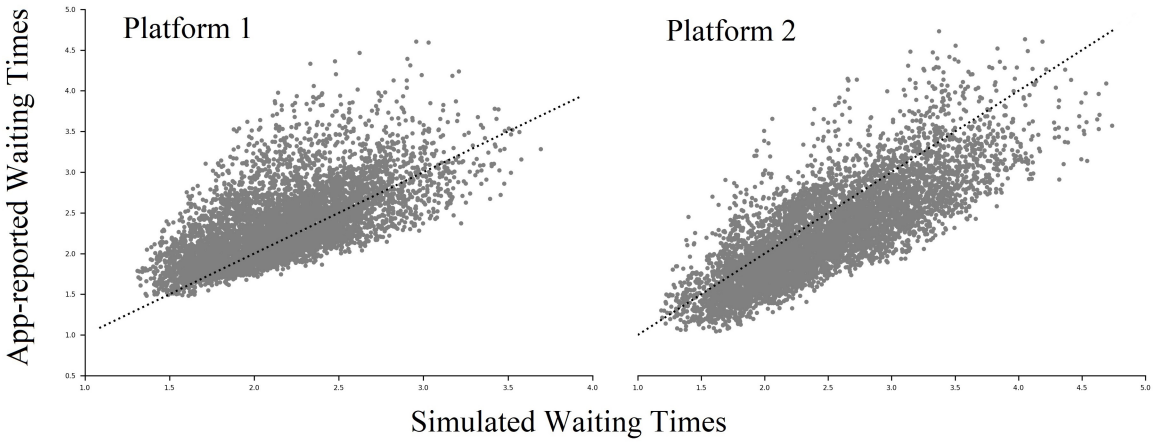

Figure 8: Comparison between the simulated waiting times and the expected waiting times reported by platforms (45 degree lines). Each point represents the average waiting time at the weekday-hour level in a particular region.

3.4 Drivers' Multi-Homing

In two-sided markets, the term multi-homing usually refers to the possibility for users of joining multiple platforms. In the current context, on the one hand, a substantial number of drivers work for both platforms, periodically switching the platform they work for. On the other hand, most drivers are not shared, meaning that they do not work for both platforms *simultaneously* (i.e. by running their driver apps concurrently). The first fact can be easily established by looking at public data sources.[20] This section provides suggestive evidence for the second.[21] These observations will later inform how I model drivers' multi-homing.

To identify shared drivers I study vehicles' GPS trajectories. If a driver is shared, one should observe two vehicles on different platforms with coincident trajectories, as depicted in Figure 9. To compare GPS trajectories across platforms, I aggregate time into windows of t seconds and space into squares of m squared meters. I represent a trajectory by means of the sequence $((s_1, w_1), ..., (s_n, w_n))$ of the squares $s_1, ..., s_n$ visited by the vehicle during the time windows $w_1, ..., w_n$. I consider two trajectories as coincident if their intersection covers at least a fraction ϵ of the shortest one. t, m and ϵ are the parameters of the

[20] According the the monthly industry reports published by the TLC, in June 2019 the two platforms dispatched about 76500 and 52000 unique vehicles, repectively, while all platforms combined dispatched about 85500 vehicles in total. This implies that at least 43000 vehicles were dispatched by both platforms, about 50% of the total fleet.

[21] Jiang et al. [2018] performed a similar analysis, reaching similar conclusions.

test, determining the degree of similarity between trajectories which is required for them to be considered coincident.

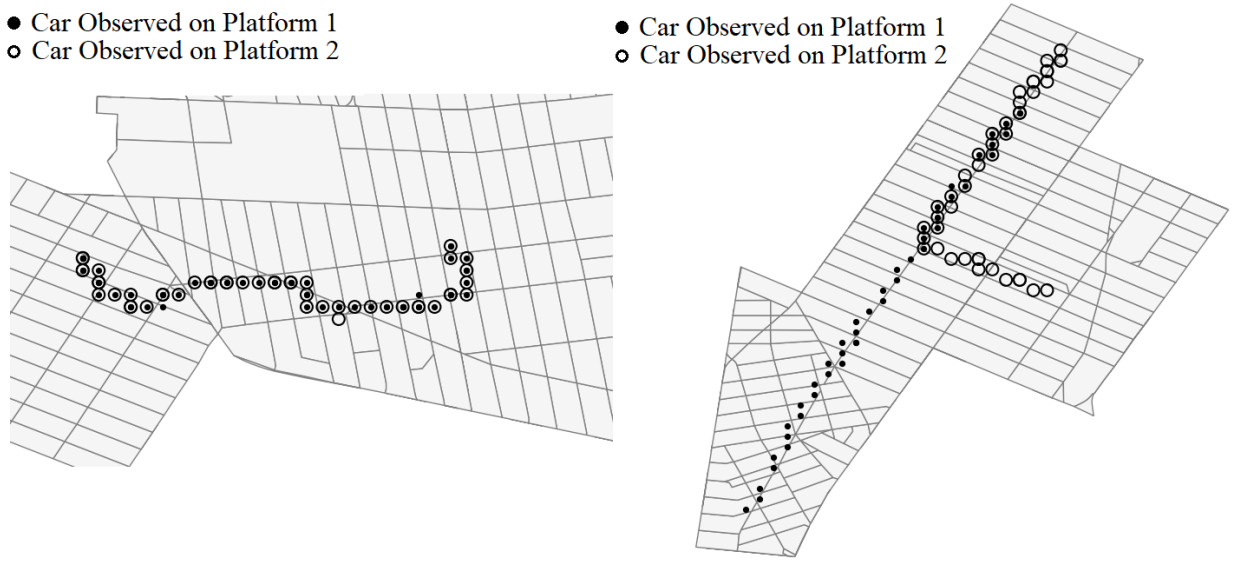

Figure 9: Two pairs of GPS trajectories observed on different platforms. On the left panel the trajectories are coincident, suggesting that one driver is working simultaneously on both platforms. On the right panel they are not, suggesting that they belong to distinct drivers.

I experiment with different values of t, m, ϵ for a one-week period, computing the share of coincident trajectories between vehicles observed on different platforms. The results, reported in Table 1, should be considered an upper bound.[22] The results suggest that the greatest majority of drivers are active on a single platform at a time. This could be due to several reasons, such as technical challenges in operating both apps at the same time, and to the fact that platforms offer promotions that disincentivize multi-homing.[23]

[22] I consider the results as upper bounds since I allow for double counting, meaning that a trajectory on one platform can be matched with multiple trajectories on the other.

[23] Both platforms have promotions that incentivize drivers to make consecutive trips, by offering extra money for completing a series of trips without canceling, rejecting, or going offline. On the technical side, operating both platforms simultaneously requires drivers to log out from one app whenever they receive a trip request from the other, and log in again once they complete the trip. Moreover, these apps communicate the driver's position to the server at short intervals, making an intensive use of data, which might require the driver to operate them on separate phones and to purchase larger data plans.

m	t	ϵ	% Shared Drivers
50	10	.8	4%
50	10	.9	1.7%
50	30	.8	5.2%
50	30	.9	2.4%
50	60	.8	5.3%
50	60	.9	2.4%
100	10	.8	7.2%
100	10	.9	4%
100	30	.8	8.7%
100	30	.9	5.4%
100	60	.8	8.7%
100	60	.9	5.5%

Table 1: The results of the test for shared drivers. Higher values of m, t and lower values of ϵ correspond to lower degree of similarity between GPS trajectories required for them to be considered the same driver.

4 Model

4.1 Model Overview

I model a transport economy populated by two types of agents, drivers and passengers. A finite set of companies $m \in M$ compete to attract agents on their platforms (marketplaces). For a generic m, I will use the notation $-m \equiv M \setminus \{m\}$ to denote the set of m's competitors.

Time and Space Time is continuous, denoted by t, and measured in minutes. There is a finite set of regions denoted by $i \in I$, and each pair of regions $ij \in I^2$ is connected by a single route.

Demand Potential passengers, also referred to as customers, can request trips on different platforms, specifying a desired route. Demand for trips on each route is a function of prices, expected waiting times, and an aggregate demand shifter $\xi(t)$ following an exogenous Markov process, which is meant to capture the intra-day variation in travel patterns documented in Section 3.

Drivers There is a large pool of potential drivers who can choose whether to hold a driving license. This decision captures the *extensive margin* of labor supply, determining the number of *licensed drivers*

in the economy. Licensed drivers can work for a single platform at a time. They can enter, exit, switch platforms, and re-locate across regions. In particular, they solve a dynamic problem where they face two types of decisions: entry/exit and movement decisions. The former capture the *intensive margin* of labor supply, modeling how often drivers work for different platforms, conditional on holding a license. The latter capture their search behavior, modeling how idle drivers relocate across regions while they wait for trip requests. When making entry/exit decisions, licensed drivers trade-off the expected profits of working on different platforms with a flow opportunity cost of time $\nu(t)$. ν plays the role of an *aggregate supply shifter,* following an exogenous Markov process which is meant to capture the intra-day variation in drivers' utility from leisure and wages associated with their outside options. The expected profits of working on different platforms are a function of drivers' compensations for trips on different routes, and of the rate at which they expect to receive trip requests in different regions.

To model the extensive margin, that is, potential drivers' decision of whether to hold a license, I assume that *licensed drivers do not discount future payoffs.* Intuitively, this means that they solve for a policy maximizing the expected long-run average (Cesaro mean) of their payoffs.[24] This criterion has the advantage that it allows to capture the value of the dynamic program facing drivers by means of a single number, its optimal expected average payoff flow, which can be interpreted as the amortized value of holding a driving license. This allows me to model the extensive margin by means of a simple free entry condition, where potential drivers trade off the license value with a fixed outside option, capturing the amortized licensing costs (e.g. period vehicle inspections, renewal taxes, and so on).

Pricing Platforms maximize their profits, which are given by the difference between revenues from passenger prices and costs from driver compensations. They do so by changing prices and compensations depending on the conditions of demand and supply, which are captured by the demand and supply shifters $\xi(t)$ and $\nu(t)$.

Information In order to make decisions agents form expectations about match qualities and prices/compensations. I assume that $\xi(t), \nu(t)$ is public information, and that agents form expectations conditional on it. So, in particular, they are able to infer the distribution of prices and compensations in the network. $\xi(t), \nu(t)$

[24] The full scope of this assumption is discussed in depth in Section 4.10.

can be thought as changing by hour of the day, in which case it simply captures the intra-day variation in travel patterns and drivers' utility from leisure. Otherwise it can be thought as capturing other exogenous factors shifting supply and demand, such as weather conditions. The idea is that endogenous variables such as prices, trip requests and the distribution of vehicles in the network co-move with $\xi(t), \nu(t)$, depending on pricing strategies and agents' behavior. Hence $\xi(t), \nu(t)$ provides information about the economic conditions, so that agents can form expectations of the payoff-relevant variables conditional on it.

Equilibrium Platforms *can avoid coordination failures* by committing to achieve given match qualities, facing a truthfulness constraint. This constraint requires the commitment to be consistent with the expectations that agents can form based on their information. At equilibrium, all platforms' commitments and pricing strategies are optimal, given those of the competitors. The equilibrium pins down a dynamic distribution of demand and supply in the network. Moreover, on the supply side, it pins down the number of licensed drivers, and the distribution of potential entrants (licensed drivers that are currently inactive, but potentially can start working from specific regions) in time and space.

Outline The presentation is organized as follows. I start by describing the matching process, agents' information and the pricing process. I then describe agents' optimal behavior conditional on expectations and prices, and the associated long-run distribution of drivers. I then define the Transport Equilibrium, which describes the equilibrium of the passenger-driver economy for any given profile of platforms' pricing strategies. Finally, I describe platforms' optimizing behavior, characterize the profit-maximizing allocations and prices, and describe the full equilibrium of the model. I conclude the section with a detailed discussion of the main assumptions of the model.

4.2 Matching Process

At time t, trip requests on route ij for platform m arrive at Poisson rate $q_{ij}^m(t)$, while the number of drivers working for m available at i is denoted by $s_i^m(t)$. q and s are determined endogenously by agents' optimal behavior, as described later. Platforms instantaneously match trip requests with drivers available at origin. Their matching technology is described by a convex and differentiable function $\boldsymbol{w}_i(s_i^m(t))$, capturing the relationship between car supply and waiting time as in Section 3.3. If a trip request arrives on ij, m, a

random driver available at i on m is matched instantly. Hence drivers available at i on m receive trip requests to j at the endogenous *matching rates* $q_{ij}^m(t)/s_i^m(t)$. In what follows, I will refer to passenger waiting times and driver matching rates as *match qualities*, since these affect the attractiveness of joining a platform for users on both sides.

4.3 Information

The evolution of the demand and supply primitives is captured by the exogenous Markov process

$$x(t) = (\xi(t), \nu(t))$$

of demand and supply shifters. I assume that $x(t)$ is *irreducible* and it can assume a *finite* set of values. Abusing notation, I denote these values by $x \in X$, so that x denotes both the process and its realizations. In what follows I refer to x as "the state", and denote by $\xi(x)$ and $\nu(x)$ its demand and supply components. Its evolution is governed by the exogenous transition rates

$$\chi = (\chi_{x,x'})_{x,x' \in X}$$

where $\chi_{x,x'}$ denotes the rate at which it jumps from x to x'. For each x, I will denote by $\chi_x \equiv \sum_{x'} \chi_{x,x'}$ the rate at which the state jumps. Given irreducibility, χ admits a *unique stationary distribution* μ over X such that, for all states x,

$$\mu(x)\chi_x = \sum_{x'} \mu(x')\chi_{x',x}$$

I assume that $x(t)$ is *public information* at time t. I also assume that it is the only source of information for agents, that is, at time t, *agents only know* $x(t)$. Conditional on this information, agents form expectations about the match qualities prevailing on different platforms. The *expected match qualities* on platform m are represented by a system $\mathcal{E}^m = (w^m, \theta^m)$ where, for every x, ij, $w_i^m(x)$ and $\theta_{ij}^m(x)$ denote the expected passenger waiting time and the expected driver matching rate conditional on x, respectively. In equilibrium, these expectations are required to be consistent with the true market dynamics. The profile of expected match qualities on all platforms will be denoted by $\mathcal{E} = (\mathcal{E}_m)_{m \in M}$.

4.4 Pricing

Platforms change prices and compensations depending on the demand and supply primitives, which are captured by the state x. Hence prices are measurable with respect to the state, so that agents' information is sufficient to predict them. A *two-sided price system* on platform m is denoted by $\mathcal{P}^m = (p^m, r^m)$ where, for every x, ij, $p^m_{ij}(x)$ and $r^m_{ij}(x)$ denote the price paid by passengers and the compensation received by drivers for trips on ij, m in state x, respectively. The profile of two-sided price systems on all platforms will be denoted by $\mathcal{P} = (\mathcal{P}_m)_{m \in M}$.

4.5 Demand

In state x, trip requests on ij, m arrive at the endogenous rate

$$q^m_{ij}(x) = \boldsymbol{q}^m_{ij}(p_{ij}(x), w_i(x), \xi(x)).$$

In particular, the demand rates depend on time only through the state, since prices and expected waiting times are measurable with respect to it. For all ij, $\boldsymbol{q}_{ij}$ denotes the demand curve for trips on ij, a function of prices, expected waiting times and the demand shifter. It is kept general for now, while in the empirical specification in Section 5.1 it will be derived from a nested logit choice model of transport mode.

4.6 Supply

This Section describes how drivers' behavior determines the long-run dynamic distribution of car supply in the network. I start by taking the number of drivers in the economy (the number of licensed drivers) as given, describe the dynamic problem they face under a given process of compensations and expected matching rates, and how their optimal behavior determines the intensive margin of supply. I later make the number of licensed drivers in the economy endogenous, thus modeling the extensive margin as well. In order to do so, I assume that drivers do not discount future payoffs, which allows me to model the extensive margin by means of a simple free entry condition. For those unfamiliar with undiscounted dynamic problems, in order to provide some intuition, I start by describing drivers' problem with discounting, and derive the optimality conditions for patient drivers by taking the limit of the Bellman equations as the

discount rate goes to zero. I then describe the long-run frequency at which drivers visit different regions associated with the optimal solution, which will later be used to define rationality of agents' expectations. I conclude the section by presenting the free entry condition.

4.6.1 Licensed Drivers with Discounting

Consider a licensed driver with discount rate ρ. At a given instant, he can be engaged in one of the following activities: working for platform m, searching at some region i; working for platform m, transporting a passenger on some route ij; inactive (potential entrant) at some region i. The driver's *individual state* is given by the combination of the exogenous state x, and the activity he is engaged in. In what follows, I refer to drivers' individual states as "*searching at* x, i, m", "*traveling on* x, ij, m", and "*inactive at* x, i". There are two types of decisions that the driver makes: entry/exit and direction decisions. Start by considering a driver facing the entry/exit decision at i, x. The inclusive value of this choice is

$$V_i(x) = \mathbb{E} \max_{m \in M \cup \{out\}} [W_i^m(x) + \frac{\eta_m}{\rho + \chi_x}] \tag{1}$$

Where η is an i.i.d. preference shock with continuous density and full support, and $\mathbb{E}$ is the expectation with respect to its realization.[25] In words, the driver can choose whether to: i) stay inactive, in which case he gets $\eta_{out}/(\rho + \chi_x)$ and becomes inactive at x, i, whose continuation value is $W_i^{out}(x)$, or ii) start working for some platform m, in which case he gets $\eta_m/(\rho + \chi_x)$ and starts searching at x, i, m, whose continuation value is $W_i^m(x)$. I assume that *all drivers in the economy face the entry/exit choice simultaneously* whenever the state jumps, which happens at the exogenous rates χ.

A driver inactive at x, i simply waits until the opportunity to enter arrives, hence the value of being inactive satisfies the Bellman equation

$$\rho W_i^{out}(x) = \sum_{x'} \chi_{x,x'} [V_i(x') - W_i^{out}(x)]. \tag{2}$$

[25] Formally, η_m is interpreted as a shock to the flow payoff of working for platform m (or being inactive, if $m = out$). The driver receives it until he faces the next entry/ext choice, which happens at rate χ_x. Hence the present discounted value of the flow payoffs due to the shock is given by

$$\int_0^\infty \eta_m e^{-(\rho+\chi_x)t} \mathrm{d}t = \eta_m/(\rho + \chi_x)$$

When a driver starts searching at x, i, m, he chooses a direction of motion (heading) h from a finite set H.[26] He then starts *"searching at x, i, m, headed towards h"*: while moving, he is available to receive trip requests from m originating at i. He then starts moving, paying the flow cost of time $\nu(x)$ until either of the following events occur:

- At rate $\theta_{ij}^m(x)$, he receives a trip request to a destination j and starts traveling on x, ij, m. The rate at which the driver gets matched to some destination is then given by

$$\theta_i^m(x) \equiv \sum_j \theta_{ij}^m(x)$$

- At an exogenous rate $\tau_{h,ij}$, which depends on where he is heading, he relocates to an adjacent region j, and faces the direction decision there.

- At rate $\lambda - \sum_j \tau_{h,ij}$ he can change direction, facing again the direction choice at i. λ is an exogenous parameter, capturing the frequency at which drivers re-optimize the direction of movement.

- At the exogenous rate $\chi_{x,x'}$, the state jumps to x' and the driver faces the entry/exit choice at x', i.

Intuitively, the driver is allowed to move while he waits for trip requests, and the direction in which he moves affects the probability according to which he reaches different nearby regions. The inclusive value of the direction choice is given by

$$W_i^m(x) = \mathbb{E} \max_{h \in H} [U_{i,h}^m(x) + \frac{\epsilon_h}{\rho + \theta_i^m(x) + \lambda + \chi_x}] \tag{3}$$

where ϵ is an i.i.d. taste shock with continuous density and full support, and $\mathbb{E}$ is the expectation with respect to its realization.[27] $U_{i,h}^m(x)$ is the value of searching at x, i, m, headed towards h, which satisfies

[26] In the empirical specification I let $H = \{\text{North}, \text{South}, \text{East}, \text{West}, \varnothing\}$, where $\varnothing$ represents the choice of waiting, i.e. not moving in any direction.

[27] Similarly to the case of the entry/exit choice, ϵ_h is interpreted as a shock to the flow payoff of searching headed towards h. The driver receives it until he either gets matched, re-optimizes the direction of movement, or faces the entry/exit choice. Either of these events occurs at rate $\theta_i^m(x) + \lambda + \chi_x$. Hence the present discounted value of the flow payoffs due to the shock is given by

$$\int_0^\infty \epsilon_h e^{-(\rho + \theta_i^m(x) + \lambda + \chi_x)t} \mathrm{d}t = \epsilon_h / (\rho + \theta_i^m(x) + \lambda + \chi_x)$$

the Bellman equation

$$\rho U^m_{i,h}(x) = -\nu(x) + \sum_j \theta^m_{ij}(x)[r^m_{ij}(x) - c^d d_{ij} + J^m_{ij}(x) - U^m_{i,h}(x)]$$
$$+ \sum_j \tau_{h,ij}[W^m_j(x) - U^m_{i,h}(x)] + (\lambda - \sum_j \tau_{h,ij})[W^m_i(x) - U^m_{i,h}(x)] + \sum_{x'} \chi_{x,x'}[V_i(x') - U^m_{i,h}(x)] \quad (4)$$

If the driver receives a trip request towards j he receives the compensation $r^m_{ij}(x)$, and pay a lump-sum travel cost $c^d d_{ij}$, which is a function of the distance d_{ij} between origin and destination and a cost c^d per unit of distance. He then starts traveling on x, ij, m, whose continuation value is $J^m_{ij}(x)$, which satisfies

$$(1 + \rho t_{ij}) J^m_{ij}(x) = -t_{ij}\nu(x) + \sum_{x'} t_{ij}\chi_{x,x'} V_j(x') + (1 - t_{ij}\chi_x) W^m_j(x) \quad (5)$$

where t_{ij} denotes the travel time between i and j. In words, if a driver gets matched he travels to destination, paying the cost of time $\nu(x)$ for t_{ij} time units. At arrival there are two possibilities: with probability $t_{ij}\chi_{x,x'}$ the state jumps to some $x' \in X$, in which case he faces the entry/exit choice at x', j, otherwise he starts searching at x, j, m.[28]

4.6.2 No-Discounting Limit

Equations 1-5 define the solution of the dynamic problem of licensed drivers under discount rate ρ. The undiscounted solution can be defined analogously by means of Bellman-like equations. In this section I show how these can be derived by taking the limit of the Bellman equations with discounting as $\rho \to 0$. I am not going to provide formal proofs, referring the reader to Rosaia [2020] for more details. For an exhaustive treatment of undiscounted dynamic problems, see also Bertsekas [1995].

Start by considering the equation defining the value of being inactive at x, i (Equation 2), making the

[28] Formally, the travel time of a trip as stochastic. A driver keeps traveling on ij until either of the following events occur: he arrives at destination, or the state jumps to some $x' \in X$. The first event occurs at rate $1/t_{ij} - \chi_x$, in which case the driver starts searching at x, j, m. The second event occurs at rate χ_x. In this case, all traveling drivers simoultaneously face the entry/exit choice at destination. The value of travelling on ij must then satisfy the Bellman equation

$$\rho J^m_{ij}(x) = -\nu(x) + \sum_{x'} \chi_{x,x'}[V_j(x') - J^m_{ij}(x)] + (\frac{1}{t_{ij}} - \chi_x)[W^m_j(x) - J^m_{ij}(x)].$$

Rearranging terms yelds Equation 5.

dependence of the valuation on the discount rate explicit:

$$\rho W_i^{out}(x|\rho) = \sum_{x'} \chi_{x,x'}[V_i(x'|\rho) - W_i^{out}(x|\rho)]$$

The term in the left hand side is the expected average of the stream of payoffs generating from the individual state "inactive at x, i", weighted according to the discount rate ρ. To take the no-discounting limit, one can pick an arbitrary valuation, for instance the value of being inactive at an arbitrary pair $i_0, x_0 \in I \times X$, and write

$$\rho W_i^{out}(x|\rho) = \sum_{x'} \chi_{x,x'}\{[V_i(x'|\rho) - W_{i_0}^{out}(x_0|\rho)] - [W_i^{out}(x|\rho) - W_{i_0}^{out}(x_0|\rho)]\}.$$

It can be shown that, as $\rho \to 0$, the limits of all three terms in this equation are well defined. In particular we can write

$$V_i(x) \equiv \lim_{\rho\to 0}[V_i(x|\rho) - W_{i_0}^{out}(x_0|\rho)] \qquad \text{and} \qquad W_i^{out}(x) \equiv \lim_{\rho\to 0}[W_i^{out}(x|\rho) - W_{i_0}^{out}(x_0|\rho)].$$

Intuitively, the differences between valuations grow at similar rates as the discount rate decreases, due to the fact that the state space is "connected", meaning that every individual state can be reached in finite time from any other individual state. This also implies that the limit of the average expected payoff on the left hand side is independent of the initial state. That is, there exists a scalar u such that for every x, i and $m \in M \cup \{out\}$ we have

$$u = \lim_{\rho\to 0} \rho W_i^m(x|\rho)$$

Hence, in the limit, u defines the optimal expected average payoff flow associated with drivers' dynamic problem, which is independent of the initial state. Rearranging terms, the limit of the Bellman equation can be written as

$$W_i^{out}(x) = -\frac{u}{\chi_x} + \sum_{x'} \frac{\chi_{x,x'}}{\chi_x} V_i(x') \tag{6}$$

Intuitively, the amount of time the driver spends waiting is stochastic. On average it takes $1/\chi_x$ periods before he faces the entry/exit choice, and in the meantime he "discounts" the optimal payoff u. Conditional on the state jumping, it assumes value x' with probability $\chi_{x,x'}/\chi_x$, and the driver weights all possible

continuation values according to these probabilities.

Using a similar logic one can take the limit of Equation 4. To do this, define

$$\lambda_i^m(x) = \theta_i^m(x) + \lambda + \chi_x$$

Intuitively, the amount of time the driver spends driving headed towards h is stochastic, and $\lambda_i^m(x)$ is the rate at which "something happens". While searching the driver pays the cost of time and discounts the optimal value u. Conditional on something happening, the probabilities of different events are proportional to the respective rates, so that in the limit we have

$$\begin{aligned} U_{i,h}^m(x) = &- \frac{\nu(x)+u}{\lambda_i^m(x)} + \sum_j \frac{\theta_{ij}^m(x)}{\lambda_i^m(x)}[r_{ij}^m(x) - c^d d_{ij} + J_{ij}^m(x)] \\ &+ \sum_j \frac{\tau_{h,ij}}{\lambda_i^m(x)} W_j^m(x) + \frac{\lambda - \sum_j \tau_{h,ij}}{\lambda_i^m(x)} W_i^m(x) + \sum_{x'} \frac{\chi_{x,x'}}{\lambda_i^m(x)} V_i(x') \end{aligned} \tag{7}$$

Finally, the limits of Equations 5, 1 and 3 can be written as

$$J_{ij}^m(x) = -t_{ij}(\nu(x)+u) + t_{ij}\sum_{x'} \chi_{x,x'} V_j(x') + (1 - t_{ij}\chi_x) W_j^m(x) \tag{8}$$

$$V_i(x) = \mathbb{E} \max_{m \in M \cup \{out\}} [W_i^m(x) + \frac{\eta_m}{\chi_x}] \tag{9}$$

$$W_i^m(x) = \mathbb{E} \max_{h \in H} [U_{i,h}^m(x) + \frac{\epsilon_h}{\theta_i^m(x) + \lambda + \chi_x}] \tag{10}$$

Taken together, Equations 6-8 define the solution of the dynamic problem with no discounting, as a function of drivers' compensations r and expected matching rates θ on different platforms. I make this dependence explicit, denoting by

$$u \equiv \boldsymbol{u}(r, \theta)$$

the *optimal expected average payoff* associated with the dynamic problem under r, θ.

Drivers' optimal behavior is captured by the probabilities by which drivers' choose to entry, exit and move in different directions. I denote by $\sigma = \boldsymbol{\sigma}(r, \theta)$ the optimal system of choice probabilities defined by

$$\sigma_i^m(x) = \Pr\{[W_i^m(x) + \frac{\eta_m}{\chi}] = \max_{m' \in M \cup \{out\}} [W_i^{m'}(x) + \frac{\eta_{m'}}{\chi}]\} \quad m \in M \cup \{out\} \tag{11}$$

and

$$\sigma_i^{h|m}(x) = \Pr\{[U_{i,h}^m(x) + \frac{\epsilon_h}{\lambda_i^m(x)}] = \max_{h' \in H} [U_{i,h'}^m(x) + \frac{\epsilon_{h'}}{\lambda_i^m(x)}]\} \quad h \in H,\, m \in M \tag{12}$$

Given the assumption that the idiosyncratic shocks η and ϵ have full support, it is easy to see that the optimal choice probabilities $\boldsymbol{\sigma}(r, \theta)$ have also full support, as defined below.

Definition 1. A system σ of conditional choice probabilities has *full support* if it gives strictly positive weight to all possible choices, that is, for every x, i, m, h:

$$\sigma_i^m(x) > 0 \qquad \text{and} \qquad \sigma_i^{h|m}(x) > 0$$

Aggregating the behavior of all licensed drivers defines the evolution of supply in the network. The next section describes the long-run distribution associated with this process.

4.6.3 Long-Run Distribution

Licensed drivers visit different individual states according to an endogenous Markov process which depends on the matching rates θ and the choice probabilities σ. The frequency according to which they visit different states in the long-run is described by a stationary distribution π of this process. For every x, ij, m, let $\pi(s_i^m, x)$, $\pi(m_{ij}^m, x)$ and $\pi(e_i, x)$ denote the probabilities that a licensed driver is searching at x, i, m, traveling at x, ij, m and inactive at x, i in the long run, respectively. Their marginals over x must be given by μ, the stationary distribution of the exogenous state. That is, for every x we must have

$$\sum_{m,i} \pi(s_i^m, x) + \sum_{m,ij} \pi(m_{ij}^m, x) + \sum_i \pi(e_i, x) = \mu(x) \tag{13}$$

Moreover, they must satisfy the following stationarity constraints:

$$\pi(m_{ij}^m, x)\frac{1}{t_{ij}} = \pi(s_i^m, x)\theta_{ij}^m(x) \tag{14}$$

$$\pi(e_i, x)\chi_x = \sum_{x'}[\sum_m \pi(s_i^m, x') + \sum_{m,j} \pi(m_{ji}^m, x') + \pi(e_i, x')]\chi_{x',x}\sigma_i^{out}(x)$$

$$\pi(s_i^m, x)[\sum_j \theta_{ij}^m(x) + \sum_{h,j} \sigma_i^{h|m}(x)\tau_{h,ij} + \chi_x] = \sum_j \pi(s_j^m, x) \sum_h \sigma_j^{h|m}(x)\tau_{h,ji}$$
$$+ \sum_j \pi(m_{ji}^m, x)(\frac{1}{t_{ji}} - \chi_x) + \sum_{x'}[\sum_m \pi(s_i^m, x') + \sum_{m,j} \pi(m_{ji}^m, x') + \pi(e_i, x')]\chi_{x',x}\sigma_i^m(x).$$

Intuitively, these constraints require the long-run rates at which drivers exit and enter different individual states to balance. The first constraint applies this logic to the state "traveling on x, ij, m". The left hand side is the rate at which drivers exit from this state, given by the probability of being traveling times the arrival rate. The right hand side is the rate at which drivers get matched, given by the probability of being searching times the matching rate. The second constraint applies the same logic to the individual state "inactive at x, i". The rate at which drivers exit from such state is given the probability of being inactive times the rate at which x jumps, while the rate at which drivers become inactive is given by the product between the rate at which they face the entry/exit choice times the probability according to which they choose to be inactive. Finally, the last constraint applies this to the individual state "searching at x, i, m". Drivers exit from such state if either they receive a trip request, they move to an adjacent region, or x jumps. On the other hand, drivers can start searching either by relocating from a different region or by facing the entry/exit choice at x, i and choosing to work for m.

It can be shown that these probabilities are unique, as stated below.

Proposition 1. *If σ has full support then there is a unique stationary distribution $\pi \equiv \boldsymbol{\pi}(\sigma, \theta)$ satisfying constraints 13 and 14.*

Proof. Given my previous assumption that χ is irreducible, if σ has full support then drivers' Markov process is irreducible as well, hence it admits a unique stationary distribution. □

The *long-run supply distribution* $\bar{\boldsymbol{s}}(\sigma, \theta, N)$ associated with σ, θ, N is obtained by aggregating $\boldsymbol{\pi}(\sigma, \theta)$

across N licensed drivers:

$$\forall x, ij, m : \bar{\boldsymbol{s}}_i^m(x|\sigma, \theta, N) \equiv N \frac{\boldsymbol{\pi}(s_i^m, x; \sigma, \theta)}{\mu(x)} \tag{15}$$

For every x, i, m, $\bar{\boldsymbol{s}}_i^m(x|\sigma, \theta, N)$ is the expected number of drivers searching at i, m conditional on the state being x, consistent with the behavior of N licensed drivers under σ. Slightly abusing notation, in what follows I will denote by

$$\bar{\boldsymbol{s}}(r, \theta, N) \equiv \bar{\boldsymbol{s}}(\boldsymbol{\sigma}(\theta, r), \theta, N)$$

the distribution associated with drivers' optimal choice probabilities under θ, r. Function $\bar{\boldsymbol{s}}(r, \theta, N)$ can be interpreted as the supply curve for this economy, under a fixed N. Analogously, we can define the long-run distribution of potential entrants $\bar{\boldsymbol{e}}(\sigma, \theta, N)$ associated with σ, θ, N:

$$\forall x, ij, m : \bar{\boldsymbol{e}}_i(x|\sigma, \theta, N) \equiv N \frac{\boldsymbol{\pi}(e_i, x; \sigma, \theta)}{\mu(x)}$$

For every x, i, $\bar{\boldsymbol{e}}_i^m(x|\sigma, \theta, N)$ is the expected number of drivers inactive at i conditional on the state being x, consistent with the behavior of N licensed drivers under σ. It is worth stressing the difference of this framework with respect to standard entry and exit models, where the measure of potential entrants in different states is usually taken as exogenous. In contrast, by endogenizing the measure of potential entrants in different regions, the model can better capture the geographical constraints imposed by the urban network.

4.6.4 Free Entry

The advantage of the no-discounting criterion is that the value of drivers' dynamic problem is captured by $\boldsymbol{u}(r, \theta)$, which does not depend on drivers' initial state. This allows to model the extensive margin of labor supply by imposing a simple *free entry condition*

$$\boldsymbol{u}(r, \theta) = \bar{u}.$$

$\bar{u}$ is a fixed outside option, interpreted as the amortized cost of holding a license. It captures costs associated to maintaining a vehicle, periodic inspections from regulatory agencies, and in the counterfactuals

it will be used to model a tax on driving licenses.

It is worth stressing once more that this framework allows to separate the intensive and extensive margin of labor supply, which are captured by licensed drivers' entry/exit decisions and by the free entry condition, respectively. In the current environment, this is important since drivers are shared among platforms on the extensive margin (licensed drivers can switch the platform they are working for) but they are not on the intensive one (licensed drivers can work for a single platform at a time).

4.7 Transport Equilibrium

In equilibrium, the expected match qualities must be consistent with the endogenous market dynamics. Intuitively, agents can compute the long-run frequency of different realizations of demand and supply conditional on realizations of the information state. Consistency requires these long-run frequencies to be compatible with the expected match qualities.

Formally, I let an *allocation of platform* m be a pair

$$\mathcal{A}^m = (q^m, \bar{s}^m)$$

of demand rates $q_{ij}^m(x)$ an expected car supply $\bar{s}_i^m(x)$ on all platforms across the network, conditional on different realizations of x, and denote by $\mathcal{A} = (\mathcal{A}^m,\, m \in M)$ a profile of allocations on all platforms.

Definition 2. $\mathcal{E}^m$ is *consistent* with $\mathcal{A}^m$ if for all x, ij

$$q_{ij}^m(x) = \bar{s}_i^m(x)\theta_{ij}^m(x)$$
$$w_i^m(x) = \boldsymbol{w}_i(\bar{s}_i^m(x))$$

Given $\mathcal{P}, \mathcal{E}$, m's allocation is pinned down by agents' optimal behavior. Denote optimal allocations by

$$\boldsymbol{A}^m(N, \mathcal{P}, \mathcal{E}) \equiv (\boldsymbol{q}^m(p, w), \bar{\boldsymbol{s}}^m(r, \theta, N)).$$

At equilibrium, expectations are consistent with the optimal allocations, and the free entry condition must be satisfied.

Definition 3. An outcome $\mathcal{P}, \mathcal{E}, N, \mathcal{A}$ is a *transport equilibrium* if:

i) *Optimality*: For all m, $\mathcal{A}^m = \boldsymbol{A}^m(N, \mathcal{P}, \mathcal{E})$

ii) *Consistency*: For all m, $\mathcal{E}^m$ is consistent with $\mathcal{A}^m$

iii) *Free entry*: $\boldsymbol{u}(r, \theta) = \bar{u}$

Optimality requires the equilibrium allocation to be consistent with the optimal behavior of passengers and drivers. In particular, the expected supply distribution $\bar{s}$ must be consistent with the long-run frequencies resulting from drivers' dynamics as discussed in previous section.

Consistency requires the expected match qualities to be consistent with the true allocation. For drivers, it requires the expected matching rates to be consistent with their true long-run averages. For passengers, it states that they form expectations consistently with the average car supply. Notice that, if $\boldsymbol{w}$ is not linear, the expected passenger waiting times need not be equal to their long-run averages. Instead, I assume that passengers form expectations about car supply in different regions, and from these they compute the expected waiting times according to $\boldsymbol{w}$. This assumption is important since it allows me to focus on the expected values captured by $\bar{s}$, making higher moments of the ergodic distribution of drivers irrelevant for passengers. Anyway, the magnitude of this discrepancy is irrelevant in my application, since conditioning on x captures most of the variation in car supply across the network.

The free entry condition requires drivers' optimal expected average payoff to be equal to the outside option $\bar{u}$, as discussed in Section 4.6.4.

4.8 Platform Competition

Conditional on x, m makes profits at rate $\sum_{ij} q_{ij}^m(x)[p_{ij}^m(x) - r_{ij}^m(x)]$, given by the difference between revenues from passengers' prices and costs from drivers' compensations. m's long-run average profit flow can be obtained by taking the expectation according to the stationary distribution μ of x:

$$\Pi(q^m, p^m, r^m) = \mathbb{E}_\mu \sum_{ij} q_{ij}^m (p_{ij}^m - r_{ij}^m). \tag{16}$$

I assume that platforms maximize their average long-run profits in 16 in a simultaneous move game, where each platform m *commits* to a pair $\mathcal{P}^m, \mathcal{E}^m$. m's commitment must be truthful, as stated below.

Definition 4. $\mathcal{P}^m, \mathcal{E}^m$ *truthful* under $\mathcal{P}^{-m}, \mathcal{E}^{-m}$ if

i) There exists N such that $\mathcal{E}^m$ is consistent with $\boldsymbol{A}^m(N, \mathcal{P}, \mathcal{E})$

ii) $\mathcal{P}, \mathcal{E}$ satisfies the free entry condition $\boldsymbol{u}(r, \theta) = \bar{u}$

In words, m takes the two-sided prices $\mathcal{P}^{-m}$ and expected match qualities $\mathcal{E}^{-m}$ of the competitors as given. In particular, m expects the competitors to provide passengers with expected waiting times w^{-m}, and expects drivers working for the competitors to be matched according to θ^{-m}. Then, predicting agents' optimal choice, m can compute the allocation $\boldsymbol{A}^m(N, \mathcal{P}, \mathcal{E})$ associated to any $N, \mathcal{P}^m, \mathcal{E}^m$. Intuitively, this means that m knows its own residual demand and supply curves. Truthfulness requires the platform's commitment to be feasible, internalizing free entry and consistency of the expected match qualities with an optimal allocation.

Definition 5. A transport equilibrium $\mathcal{P}, \mathcal{E}, N, \mathcal{A}$ is a *platform equilibrium* if, for every m, $\mathcal{P}^m, \mathcal{E}^m$ solves

$$\begin{aligned} &\max_{\mathcal{P}^m, \mathcal{E}^m} \Pi(\boldsymbol{q}^m(p, w), p^m, r^m) \\ &\quad \text{s.t. } \mathcal{P}^m, \mathcal{E}^m \text{ is thruthful under } \mathcal{P}^{-m}, \mathcal{E}^{-m} \end{aligned} \tag{17}$$

In words, at equilibrium all platforms best respond, choosing a truthful commitment maximizing their average profits, taking competitors' commitments as given. In particular, by selecting $\mathcal{E}^m$, platforms can avoid possible *coordination failures* arising from the fact that, for any $\mathcal{P}^m$, there are multiple $\mathcal{E}^m$ consistent with $\boldsymbol{A}^m(N, \mathcal{P}, \mathcal{E})$. This feature is widespread in two-sided markets, where typically users on one side join a platform only if they expect users on the other side to join as well, hence multiple equilibria might arise depending on whether users on different sides coordinate to participate or not. Intuitively, platforms deal with this multiplicity by "selecting an equilibrium in their own marketplace". A more detailed discussion is presented is Section 4.10.

4.9 Characterization of Platforms' Best Response

This section provides a more insightful characterization of Problem 17, and more suitable for computation. In what follows, consider a generic platform m and take $\mathcal{P}^{-m}, \mathcal{E}^{-m}$ as given. The idea is to characterize the set of allocations that the platform can truthfully implement, and the profits associated to such

allocations, so that 17 can be re-stated as an allocation choice problem.

Definition 6. $\mathcal{P}^m, \mathcal{E}^m$ implements $\mathcal{A}^m$ if it is truthful and there exists N such that $\mathcal{A}^m = \boldsymbol{\mathcal{A}}^m(N, \mathcal{P}, \mathcal{E})$

On the demand side I assume that, for every x, ij, there exists a differentiable *inverse demand curve*

$$p_{ij}^m(x) = \boldsymbol{p}_{ij}^m(q_{ij}^m(x), w_i^m(x) | p_{ij}^{-m}(x), w_i^{-m}(x), \xi(x)) \tag{18}$$

whose domain is the entire $\mathbb{R}_+^2$. That is, for every expected waiting time $w_i^m(x)$ and every demand rate $q_{ij}^m(x)$, function $\boldsymbol{p}_{ij}^m$ pins down a price $p_{ij}^m(x)$ such that $q_{ij}^m(x) = \boldsymbol{q}_{ij}^m(p_{ij}(x), w_i(x), \xi(x))$. Conditions for the invertibility of demand are well known in the literature, hence I will not pursue them here. Under this assumption the demand side does not impose constraints on the allocations that platforms can implement. On the supply side, platforms face constraints imposed by the network structure. Define the set of feasible long-run allocations as follows.

Definition 7. The set of feasible long-run allocations $\mathcal{L}(\theta^{-m})$ is the set of all $q^m, \bar{s}^m > 0$ such that, defining $\theta_{ij}^m(x) = q_{ij}^m(x)/\bar{s}_i(x)$ for every x, ij, m, we have $\bar{s}^m = \bar{\boldsymbol{s}}^m(\sigma, \theta, N)$ for some system of choice probabilities σ with full support.[29]

That is, $\mathcal{L}(\theta^{-m})$ is the set of all allocations that can be supported in the long-run by some behavior of drivers, hence it captures the constraints on implementation imposed by the network structure. The next result shows that these are the only constraints that platforms face, and that the total drivers' revenue are pinned down by the allocation.

Theorem 1. *For every allocation $q^m, \bar{s}^m \in \mathcal{L}(\theta^{-m})$ there exists $\mathcal{P}^m, \mathcal{E}^m$ implementing it. Moreover, there exists a convex and differentiable function $\boldsymbol{C}^m(\cdot | r^{-m}, \theta^{-m})$ over $\mathcal{L}(\theta^{-m})$ such that, for every such $\mathcal{P}^m, \mathcal{E}^m$:*

i) Drivers' compensations satisfy

$$\mathbb{E}_\mu \sum_{ij} q_{ij}^m r_{ij}^m = \boldsymbol{C}^m(q^m, \bar{s}^m | r^{-m}, \theta^{-m}).$$

[29] Inequalities between functions are interpreted point wise, so that $q^m, \bar{s}^m > 0$ means that $q_{ij}^m(x), \bar{s}_i^m(x) > 0$ for every x, ij.

ii) Letting V, W, U be drivers' value functions under r, θ, for every x, ij we have

$$\frac{d\boldsymbol{C}^m(q^m, \bar{s}^m | r^{-m}, \theta^{-m})}{dq_{ij}^m(x)} = W_i^m(x) + c^d d_{ij} - J_{ij}^m(x) \tag{19}$$

$$\frac{d\boldsymbol{C}^m(q^m, \bar{s}^m | r^{-m}, \theta^{-m})}{d\bar{s}_i^m(x)} = \sum_j \theta_{ij}^m(x)[r_{ij}^m(x) - c^d d_{ij} + J_{ij}^m(x) - W_i^m(x)] \tag{20}$$

Function $\boldsymbol{C}^m$ can be interpreted as the technological frontier for the platform. The Appendix shows that it is the value of an optimization problem, which consists in choosing a system of drivers' choice probabilities consistent with q^m and $\bar{s}^m$ in order to minimize total cost arising from drivers' activities. This can be seen as a first welfare theorem, restricted to drivers' behavior conditional on supporting a given allocation. The relationship between $\boldsymbol{C}^m$ and drivers' costs gives rise to the equivalence between marginal costs and drivers' valuations in Equations 19 and 20. Intuitively, $\boldsymbol{C}^m$ measures drivers' total costs, hence its gradient measures the costs for a single driver. In particular its gradient with respect to q measures the dynamic cost for a driver of receiving trip requests on different routes, while its gradient with respect to $\bar{s}$ measures the dynamic cost for a driver of being searching in different regions. Compensations must fully compensate drivers' for these costs. This can be seen from Equation 20, where the left hand side is the expected dynamic cost for a driver of being searching at x, i, m, while the right hand side is his expected dynamic matching surplus. Equation 19 is also intuitive, showing that the dynamic cost for a driver of receiving a trip request trades off the missed earning opportunities at origin with the continuation value of being matched.

Theorem 1 implies that Problem 17 is equivalent to the following allocation choice problem:

$$\max_{q^m, \bar{s}^m \in \mathcal{L}(\theta^{-m})} \mathbb{E}_\mu \sum_{ij} q_{ij}^m \boldsymbol{p}_{ij}^m(q_{ij}^m, \boldsymbol{w}_i(\bar{s}_i^m) | p_{ij}^{-m}, w_i^{-m}, \xi) - \boldsymbol{C}^m(q^m, \bar{s}^m | r^{-m}, \theta^{-m}). \tag{21}$$

In words, the platform can choose the quantity and "quality" of trips on every route. The latter is captured by the expected supply in different regions, which in turn determines passengers' value for the service through the expected waiting times. That is, platforms in this model behave as (residual) Spence's monopolists, which implies that there are two sources of inefficiency. The first is the classic market power distortion, implying that quantities are always locally under provided. The second is the Spence's distortion, implying that vehicles might be either under or over provided depending on the

characteristics of passengers' demand curve. For more details, see Spence [1975].

This allows for a simple characterization of platforms' optimality conditions, which provide nice insights into platforms' price-setting behavior, as discussed in the next section. Moreover notice that, if revenues are concave, 21 is a concave maximization problem with a known gradient, hence platforms' best response can be computed through standard algorithms.

4.9.1 Optimal Prices

Taking the first order conditions with respect to $\bar{s}$ we can see that, at optimum, compensations are set so to align drivers' expected matching surpluses with the marginal revenues from quality:

$$\frac{\mathrm{d}\boldsymbol{w}_i(\bar{s}_i^m(x))}{\mathrm{d}\bar{s}_i^m}\sum_j q_{ij}^m(x)\frac{\mathrm{d}\boldsymbol{p}_{ij}^m}{\mathrm{d}w_i^m}(x) = \sum_j \theta_{ij}^m(x)[r_{ij}^m(x) - c^d d_{ij} + J_{ij}^m(x) - W_i^m(x)].$$

This is reminiscent of familiar conditions in the search and matching literature. In particular, it is well known that efficiency in search models typically requires that the matching surpluses of individual agents be equal to the marginal social value of an additional agent joining the search pool.[30] The difference is that, in this case, platforms maximize their private profits instead of social welfare, hence at optimum drivers' incentives are aligned with profit maximization.

The first order conditions with respect to q yield a closed for expression for the optimal prices in terms of drivers' value functions. Using the definition of $J_{ij}^m(x)$ in equation 8, these can be written as

$$\begin{aligned} p_{ij}^m(x) =& t_{ij}(\nu(x) + \bar{u}) + c^d d_{ij} + q_{ij}^m(x)\left|\frac{\mathrm{d}\boldsymbol{p}_{ij}^m}{\mathrm{d}q_{ij}^m}(x)\right| \\ &+ W_i^m(x) - (1 - t_{ij}\chi_x)W_j^m(x) - t_{ij}\sum_{x'}\chi_{x,x'}V_j^m(x') \end{aligned} \tag{22}$$

from which it can be seen that the optimal prices are the sum of three components: a "static" marginal cost based on time and distance, a classic market power distortion, and a dynamic component trading off the marginal profitability of drivers at origin with the marginal benefit of expanding supply at destination.

[30] In models with bargaining, this condition is equivalent to the familiar Hosios [1990] conditions. See also Brancaccio et al. [2020b] for an example related to transport markets.

4.10 Discussion

This Section discusses more in depth three critical aspects of the model: i) the assumption that drivers do not discount future payoffs, ii) the assumptions on information and pricing, and iii) the way in which platform competition in modeled.

No Discounting The no discounting assumption has two main advantages. First, as already discussed, it allows to model labor supply on the extensive margin. Second, it allows to simplify the treatment of platform optimization. Intuitively, the simplification comes from the fact that drivers and platforms only care about the distribution of their payoffs in the long-run, which allows to ignore the whole trajectory of the dynamic process, focusing directly on its ergodic distributions - however notice that the latter are not associated to steady states, since in the long-run the expected prices, supply and demand still fluctuate depending depending on the exogenous information state x. On the theoretical side, this assumption is not crucial, what is crucial instead is that drivers and platforms discount future payoffs in the same way. If the former assumption was removed but the latter was maintained, a similar treatment would be possible, but more cumbersome, due to the fact that platform maximization would become state-dependent - platforms would implement a dynamic trajectory of supply and demand over the network conditional on any initial state of demand and supply. The need of dealing with all possible trajectories would also raise challenges on the computational side, due to a curse of dimensionality.

One line of defense of the no-discounting criterion is as an approximation of models with discounting. Indeed, it can be shown that drivers' choice probabilities under no-discounting are the limit of those with discounting, and in particular there is no selection involved in taking the limit. Intuitively, this is due to the fact that drivers' idiosyncratic preference shocks have full support, ruling out "extreme points" as solutions of the undiscounted model. Another line of defense is as a good description of the true optimizing behavior of drivers. For instance, drivers might learn what is the best policy in order to maximize their average weekly earnings. In general, the no-discounting criterion is appealing in situations where agents face frequent decisions in a recurrent fashion.

Information and Pricing The model assumes that platforms set prices conditional on public information, and that the latter is limited to an exogenous information source. In the empirical specification

I take this state to be hour and weekday, and the reasons for this choice are discussed at the beginning of Section 5. To better understand this assumption it is useful to conceptually separate this information from the endogenous state, describing the distribution of demand and supply in the network. Given this, the assumption can be separated in two components: i) public information is limited to the exogenous state, and ii) prices do not reveal additional information on top of the public source. For theory purposes the first component could be relaxed in several ways. It is introduced mainly for empirical reasons, since my estimation strategy requires observing the expected prices, demand and supply conditional on all realizations of public information. To understand the second, notice that platforms might have access to additional information on top of what is observed by agents, and might condition prices on it. In this case, observing the current prices would provide drivers with additional information about the future path of compensations and matching rates, informing them about the current state of platforms' private information. In particular, current prices would provide drivers with information which is not already captured by the public source. I have not experimented generalizing the model to the case in which agents and platforms have access to different information, and I leave this to future work.

Platform Competition To better isolate the assumptions on platform competition it is useful to discuss first the monopoly case, that is, the case in which the set of platforms M is a singleton. In this case, the main assumption is that, by selecting agents' expected match qualities (consistently with rational expectations) the monopolist platform can solve coordination failures. Intuitively, coordination failures might arise when drivers do not join the platform because they expect passengers not to join, and vice-versa, and these expectations are self-fulfilled in equilibrium. This implies that, for any given pricing strategy of the monopolist, multiple equilibria might arise. Setting the expected match qualities gives the monopolist the power to select one of these equilibria. Assumptions of this this kind are common in the literature on two-sided platforms. For instance, Armstrong [2006] assumes that, instead of the two-sided prices, the platform can choose the utility levels provided to users on different sides, thus ruling out the multiplicity. Weyl [2010] makes another step, showing that the platform can select an equilibrium by committing to change prices depending on users' participation levels, thus "insulating" the utility of users different sides. In this paper I do not take the step of deriving these insulating tariffs, but instead

I assume that platforms can credibly commit to achieve given match qualities.[31] Moving to platform competition, an additional assumption is introduced: when considering whether to deviate, platforms take agents' expectations about the match qualities of the competitors as given. In particular, platforms only internalize consistency of the allocation with their own match qualities, but not the consistency constraints of the competitors. This implies that the allocation resulting from a platform's deviation might not be consistent with agents' expectations about the match qualities of the competitors. Hence, by deviating, a platform forces its competitors to deviate as well. A a platform equilibrium is a rest point of this process, where the consistency constraints are simultaneously internalized by all platforms. This is only one of many possible ways of modeling competition in this setup. For instance, one alternative would be to assume that platforms have instruments to insulate their users' participation rates against deviations of the competitors, which would be equivalent to Cournot competition in quantities and qualities. In this case it can be shown that platforms face a slightly different cost function than the one derived in Section 4.9, which takes into account the effect of a platform's deviation on the compensations set by its competitors. A more thoughtful approach would be to model explicitly the pricing instruments that platforms might use to achieve equilibrium selection, an approach that is not pursued in this paper but is left for future work.

5 Estimation

As anticipated, when I bring the model to the data I assume that the demand and supply shifters are a function of hour and weekday, so that states x are identified by different weekday-hour pairs. Formally, x jumps on average every 60 minutes, and its stationary distribution μ is uniform. Intuitively, the assumption is that demand and supply primitives (i.e. travel patterns and drivers' value of time) differ across day of the week and hour of the day, and that drivers make dynamic decisions knowing only what day and hour it is, forming correct conditional expectations about the payoff relevant variables.

The assumption on information is not literally verified in practice, where drivers have access to a variety of real-time signals through platforms' applications. However, conditioning on hour and weekday

[31] On the one hand, Weyl's tariffs could be adapted to insulate the demand side. On the supply side complexities arise due to the dynamic structure of drivers' problem, hence additional instruments might be needed.

captures most of the observed variation in the data, with different weeks displaying almost identical patterns.

5.1 Demand Estimation

To estimate demand, I use two different sources of variation. First, I identify how customers substitute between platforms by exploiting high-frequency variation in prices, waiting times, and trip requests. Second, I identify how the total number of ride-sharing trips reacts to changes in prices and waiting times by exploiting the policy change as a natural experiment, which allows me to deal with endogeneity. Formally, I estimate a nested logit discrete choice model of transport mode at the origin-destination level, whose parameters vary depending on the average trip length. I group routes in 5-minute bins depending on their average trip length, and let t_{ij} denote the bin (rounded average trip length) of route ij. At weekday-hour x, the utility of passengers facing prices p and expected waiting times w for taking a trip from i to j on platform m is given by

$$\xi_{ij}(x) + v(t_{ij}) \cdot 1\{m = a\} - \alpha(t_{ij})p_{ij}^m - \beta(t_{ij})w_i^m + \epsilon^m$$

where $1\{m = a\}$ is an indicator equal to 1 if $m = a$ and 0 otherwise. For every weekday-hour pair x, the demand shifter $\xi_{ij}(x)$ is interpreted as the average gross customer valuation for a ride-sharing trip. This valuation is unobservable to the econometrician, but perfectly observable by passengers and platforms. $v(t_{ij})$ is a "brand effect" capturing the difference between passengers' valuations for trips on platform a and trips on platform b, which depends on the average trip length t_{ij}. $\alpha(t_{ij})$ and $\beta(t_{ij})$ also vary according to the average trip length, capturing passengers' disutilities for price and waiting time, respectively.

Potential passengers considering ride-sharing travel on route ij arrive at an exogenous rate $n_{ij}(x)$. For every weekday-hour x, $n_{ij}(x)$ represents the typical number of trips per minutes on route ij, on all modes, which is estimated from the MTA travel survey as described in Section 2.2.3. Potential passengers can choose between different ride-sharing platforms $m = a, b$ and an outside option representing alternative travel modes, whose utility is given by ϵ^0.

ϵ is an idiosyncratic shock capturing variance in tastes across passengers, whose distribution is chosen so to yield the familiar nested logit shares. This distribution is parameterized by a scalar $\gamma(t_{ij})$, capturing

(one minus) the correlation between the logit shocks for alternative platforms. This is a function of the average trip length, which is to be estimated. The nested logit share of platform m is given by the product of its within market share (the share of trips on m out of all ride-sharing trips) and the market share (the share of passengers taking a ride sharing trip out of all potential passengers). These shares can be written as follows. Denote by $\tilde{v}$, $\tilde{\alpha}$ and $\tilde{\beta}$ the parameter ratios defined by

$$\tilde{v}(t_{ij}), \tilde{\alpha}(t_{ij}), \tilde{\beta}(t_{ij}) = v(t_{ij})/\gamma(t_{ij}), \alpha(t_{ij})/\gamma(t_{ij}), \beta(t_{ij})/\gamma(t_{ij})$$

for every t_{ij}. The within market share of platform m can be written as

$$P_{ij}(m|M; p, w; \tilde{v}, \tilde{\alpha}, \tilde{\beta}) \equiv \frac{\exp[\tilde{v}(t_{ij}) \cdot 1\{m = a\} - \tilde{\alpha}(t_{ij}) p_{ij}^m - \tilde{\beta}(t_{ij}) w_i^m]}{\sum_{m'} \exp[\tilde{v}(t_{ij}) \cdot 1\{m' = a\} - \tilde{\alpha}(t_{ij}) p_{ij}^{m'} - \tilde{\beta}(t_{ij}) w_i^{m'}]} \tag{23}$$

whereas the market share can be written as

$$P_{ij}(M; V; \gamma, \xi(x)) \equiv \frac{\exp[\xi_{ij}(x) + \gamma(t_{ij}) V_{ij}]}{1 + \exp[\xi_{ij}(x) + \gamma(t_{ij}) V_{ij}]} \tag{24}$$

V_{ij}, defined below, can be seen as the inclusive value of the within market choice

$$V_{ij} = \log \sum_m \exp[\tilde{v}(t_{ij}) \cdot 1\{m = a\} - \tilde{\alpha}(t_{ij}) p_{ij}^m - \tilde{\beta}(t_{ij}) w_i^m] \tag{25}$$

Intuitively, the parameter ratios $\tilde{v}$, $\tilde{\alpha}$ and $\tilde{\beta}$ govern how passengers substitute between platforms, trading off prices and waiting times, while parameter γ measures the sensitivity of total ride-sharing demand to the inclusive value. The unobservable demand shifter ξ_{ij} can be used to formalize the endogeneity problem in this model: it is correlated with prices, since platforms can condition prices on it; it is also correlated with waiting times, since it determines demand for trips thus affecting car supply at origin. Notice that ξ_{ij} does not depend on m, hence it enters the outer layer in Equation 24, but it does not affect the inner layer in Equation 23. In words, *I assume that unobservable factors might affect total ride-sharing demand, but not how passengers substitute between platforms.* This allows me to estimate the model in two stages. In the first stage, I estimate the inner layer via maximum likelihood on the high-frequency data. In this way I can exploit high-frequency variation to estimate how customers substitute between platforms, and

how they trade off prices and waiting times. In the second stage, I estimate the outer layer exploiting the variation in the pre-post policy data, thus dealing with endogeneity.

5.1.1 Inner Layer

In the first stage, I estimate the parameter ratios $\tilde{v}$, $\tilde{\alpha}$ and $\tilde{\beta}$ entering the inner layer. I do so simply by running maximum likelihood on the high-frequency data described in Section 2.2.1. Recall that the sample period is discretized into 10-minute intervals, and that for each interval t I observe the average prices $p_{ij}^m(t)$, reported waiting times $w_i^m(t)$, and number of trip requests $q_{ij}^m(t)$ on every platform m and origin-destination pair ij. This allows me to compute the log-likelihood

$$l(\tilde{v}, \tilde{\alpha}, \tilde{\beta}) = \sum_t \sum_{ij,m} q_{ij}^m(t) \log P_{ij}(m|M; p(t), w(t); \tilde{v}, \tilde{\alpha}, \tilde{\beta})$$

associated to each candidate set of parameter ratios, and maximize it to obtain consistent estimates.

5.1.2 Outer Layer

To estimate the outer layer, I exploit the pre-post policy data described in Section 2.2.2. Recall that this data consists in the pre-post policy estimates of prices, waiting times and trip requests at the weekday-hour level for every platform and origin-destination pair. From these, I can use the parameter ratios estimated in the first stage to obtain consistent estimates $V_{ij}^{pre}(x)$ and $V_{ij}^{post}(x)$ of the inclusive values defined in Equation 25, before and after the policy. Also, using arrival the rates of potential passengers estimated from the MTA survey, I can obtain estimates of the market shares before and after the policy.

Given this notice that, for $k = pre, post$, 24 can be written as a linear equation

$$y_{ij}^k(x) = \gamma(t_{ij}) V_{ij}^k(x) + \xi_{ij}^k(x)$$

where the dependent variable is the log of the ratio between the market share and the share of the outside option, and all variables except $\xi_{ij}^k(x)$ are observed.[32] To estimate γ I impose that, for every average trip

[32] Formally, for $k = pre, post$, the dependent variable is defined as $y_{ij}^k(x) \equiv \log P_{ij}^k(M; x) - \log[1 - P_{ij}^k(M; x)]$.

length l, we have

$$\mathbb{E}[\xi_{ij}^{post}(x) - \xi_{ij}^{pre}(x)|t_{ij} = l] = 0$$

where the expectation is with respect to both the weekday-hour pair x and the route ij. In words the exclusion restriction is that, although there might be differences in unobservable factors affecting demand before and after the policy, these must average out across routes and weekday-hour pairs. Under this assumption I can obtain consistent estimates of γ simply by comparing pre-post policy averages

$$\gamma(l) = \frac{\mathbb{E}[y_{ij}^{post}(x) - y_{ij}^{pre}(x)|t_{ij} = l]}{\mathbb{E}[V_{ij}^{post}(x) - V_{ij}^{pre}(x)|t_{ij} = l]}$$

for every average trip length l. Finally, given the estimates of all parameters, I can recover the average customer trip valuations represented by ξ.

5.1.3 Estimates

Table 2 shows the estimates of v, α, β and γ as a function of the average trip length. The demand sensitivity to both prices and waiting times is lower for longer trips. This is intuitive, since longer trips are associated with higher prices and higher travel times on average. This implies that, for longer trips, a marginal increase in price is associated to a lower percentage increase in the total cost, and a marginal increase in waiting time is associated to a lower percentage increase in the total time of the trip (including both travel and waiting time).

Average Trip Length	v	α	β	γ
$<$ 15 min	0.776 (0.055)	0.028 (0.002)	0.218 (0.015)	0.762 (0.051)
15 min - 20 min	0.659 (0.055)	0.020 (0.001)	0.136 (0.011)	0.648 (0.045)
20 min - 25 min	0.614 (0.035)	0.019 (0.001)	0.119 (0.006)	0.604 (0.031)
25 min - 30 min	0.590 (0.033)	0.020 (0.001)	0.090 (0.005)	0.587 (0.029)
30 min - 35 min	0.539 (0.035)	0.016 (0.001)	0.013 (0.000)	0.530 (0.028)
35 min - 40 min	0.522 (0.058)	0.013 (0.001)	0.012 (0.001)	0.513 (0.039)
$>$ 40 min	0.467 (0.108)	0.011 (0.002)	0.011 (0.002)	0.459 (0.058)

Table 2: Nested Logit Coefficients as a function of the average trip length. v is the brand effect, α and β represent passengers' sensitivity to prices and waiting times, respectively. γ captures (one minus) the correlation between the logit shocks for different platforms. The standard errors are in parentheses.

To better compare these estimates with previous studies, Table 3 shows the associated demand elasticities. I find that price elasticities increase with the trip length, ranging from about 0.5 to about 1 for the larger platform. In contrast, waiting time elasticities decrease sharply with trip the average trip length, from a maximum value of 0.56 for trips shorter than 15 minutes to a minimum value of 0.03 for trips of 45 minutes or longer for the larger platform. The elasticities for the smaller platform follow a similar pattern but they are higher, which is a mechanical consequence of the specification of the model specification.

	Platform a		Platform b	
Average Trip Length	Price	Waiting Time	Price	Waiting Time
$<$ 15 min	0.55	0.560	0.78	0.963
15 min - 20 min	0.62	0.362	0.88	0.626
20 min - 25 min	0.82	0.317	1.21	0.552
25 min - 30 min	1.07	0.251	1.53	0.441
30 min - 35 min	1.08	0.063	1.53	0.101
35 min - 40 min	1.08	0.061	1.54	0.100
$>$ 40 min	1.09	0.028	1.56	0.050

Table 3: Average demand elastiticies with respect to prices and waiting times as a function of the average trip length for the larger platform (platform a) and the smaller platform (platform b) corresponding to the estimated nested logit parameters. The averages are weighted with respect to the number of trip requests.

One useful benchmark for price elasticities is the study by Cohen et al. [2016b], which covers several cities in the United States including New York City. They find price elasticities between .35 and and 1.02 for the main platform, which are roughly in line with my estimates. Buchholz et al. [2020] estimate the waiting times elasticities for a ride-sharing platform in Prague, finding that price elasticities are four to ten times as large as waiting-time elasticities. Hence their estimates are roughly consistent with mine, considering trips of 25 minutes or longer.

5.2 Drivers' Parameters

Drivers' parameters are identified from the number of licensed drivers, and from the observed variation in the number of idle drivers, compensations and matching rates in time and space. Formally I prove that, given the observed number of licensed drivers N and expected matching rates θ, the equilibrium long-

run distribution of idle drivers $\bar{s}$ can be inverted to recover a function of drivers' payoffs, which in turn identifies the parameters. In particular, these are identified even though the equilibrium distribution of potential entrants and *the optimal choice probabilities are not observed*. In order to understand this result, it is useful to cast it in the broader literature of estimation of dynamic discrete choice models. *While standard estimation algorithms are based on the observation of the optimal choice probabilities in different states, my estimation algorithm is based on the observation of the long-run frequencies at which agents visit a subset of all states*. Indeed, recall that in Section 4.6.3 $\bar{s}_i^m(x)$ was defined as probability according to which, conditional on state x, a driver is observed searching at region i on platform m in the long-run. In particular, observing the long-run frequencies at which drivers' are inactive in different regions is not required for identification, hence this strategy is particularly appealing in the current context.

On top of the formal identification result, I provide an algorithm that can be used to recover the identified set of payoffs from $\bar{s}$. The algorithm consists in a simple tâtonnement procedure, which is reminiscent of inversion techniques commonly used for demand estimation.

In what follows I take N, θ and r as given. I start by defining the identified set of drivers' payoff. Then I state the inversion theorem and illustrate the inversion algorithm. Finally I show how this allows to recover drivers' parameters.

5.2.1 Mean Payoffs

Define the expected payoff flow of searching or, more briefly, the *mean payoff* at x, i, m, by

$$\delta_i^m(x) = -\nu(x) + \sum_j \theta_{ij}^m(x)[r_{ij}^m(x) - \nu(x)t_{ij} - c^d d_{ij}] \tag{26}$$

and notice that drivers' value functions depend on r, ν, c^d only through δ. Intuitively, a driver searching at x, i, m expects to receive utility at rate $\delta_i^m(x)$ before facing the next choice. In particular we can re-write drivers' value of searching while moving in different directions (Equation 7) as

$$\begin{aligned} U_{i,h}^m(x) = \frac{\delta_i^m(x) - u}{\lambda_i^m(x)} + \sum_j \frac{\theta_{ij}^m(x)}{\lambda_i^m(x)}[-t_{ij}\bar{u} + t_{ij}\sum_{x'}\chi_{x,x'}V_j(x') + (1 - t_{ij}\chi_x)W_j^m(x)] \\ + \sum_j \frac{\tau_{h,ij}}{\lambda_i^m(x)}W_j^m(x) + \frac{\lambda - \sum_j \tau_{h,ij}}{\lambda_i^m(x)}W_i^m(x) + \sum_{x'}\frac{\chi_{x,x'}}{\lambda_i^m(x)}V_i(x') \end{aligned} \tag{27}$$

The solution to drivers' dynamic problem is then defined by Equations 6, 8, 9, 10, and 27, non of which involves r, ν, c^d. Intuitively, one can specify an equivalent problem where drivers do not face any cost of time and transport cost, and they do not receive compensations for specific trips, receiving instead $\delta_i^m(x)$ per unit of time they spend searching in different states, regions and platforms. The two problems are equivalent, meaning that they generate the same dynamic behavior and average payoffs. In particular, they generate the same expected long-run distribution of idle drivers and the same optimal expected average payoffs, which can be derived from the optimal solution as in Section 4.6.3. Hence we can write

$$\bar{s} = \bar{\boldsymbol{s}}(\delta) \text{ and } u = \boldsymbol{u}(\delta)$$

Given this, we can expect to identify at most δ from the observation of $\bar{s}$. Next section shows that this is the case. More formally, it shows that function $\bar{\boldsymbol{s}}$ is invertible, and provides an algorithm for inversion.

5.2.2 Inversion

In what follows denote by S the set of supply distributions that are feasible under θ. Formally, S is defined as the set of all $\bar{s} > 0$ such that there exists a system of choice probabilities σ of full support such that $\bar{s} = \bar{\boldsymbol{s}}(x|\sigma, \theta, N)$

Theorem 2. *For every $\bar{s} \in S$ there exists a unique $\delta \equiv \boldsymbol{\delta}(\bar{s})$ such that $\bar{\boldsymbol{s}}(\delta) = \bar{s}$. Moreover, function $\boldsymbol{u}(\delta)$ is strictly convex, differentiable, and such that $N\nabla \boldsymbol{u}(\delta) = \bar{s}(\delta)$. Finally, for every $\bar{s} \in S$ we have*

$$\boldsymbol{\delta}(\bar{s}) = \arg\min_{\delta}[N\boldsymbol{u}(\delta) - \sum_{x,i,m} s_i^m(x)\delta_i^m(x)] \tag{28}$$

In words, the long-run distribution of idle drivers identifies the mean payoffs, and the latter can be recovered by solving a convex minimization problem. Moreover, the characterization of the objective function's gradient allows to solve this problem by means of standard convex optimization algorithms. For instance, consider gradient descent applied to Problem 28. Given a small positive constant k, starting from any initial guess δ^0, this generates the sequence $\delta^1, \delta^2, ...\delta^n, ...$ defined by

$$\delta_i^{n+1,m}(x) = \delta_i^{n,m}(x) - k[\bar{\boldsymbol{s}}_i^m(\delta^n)(x) - \bar{s}_i^m(x)]$$

That is, gradient descent corresponds to a simple tâtonnement algorithm where $\delta_i^{n,m}(x)$ is increased (resp. decreased) if $\bar{\boldsymbol{s}}_i^m(\delta^n)(x)$ is smaller (resp. higher) than $\bar{s}_i^m(x)$. At each step, $\bar{\boldsymbol{s}}(\delta)$ can be computed by solving drivers' dynamic problem, and computing the stationary distribution associated with the optimal choice probabilities under δ. For k small enough, $\delta^n \to \boldsymbol{\delta}(\bar{s})$ as $n \to \infty$.

A reader familiar with the literature of demand estimation might notice the similarity with standard algorithms performing static demand inversion, as in Berry et al. [1995]. The logic is similar. Intuitively, in the case of demand estimation, the mean payoffs associated with different products can be recovered by tâtonnement from the observation of the frequencies according to which customers choose these products. In this case, the mean payoffs associated with different activities can be recovered by tâtonnement from the observation of the frequency at which agents perform these activities in the long-run. The key difference is that in this case the function to be inverted, i.e. $\bar{\boldsymbol{s}}(\delta)$, does not satisfy the gross substitutes property, but instead invertibility is guaranteed by convex conjugacy. See the Appendix for more details on this, and Berry et al. [2013] for the theory behind demand inversion algorithms.

5.2.3 Estimating Drivers' Parameters

To estimate drivers' parameters, I assume that the i.i.d. shocks at every decision node follow a standard logit distribution, yielding the familiar logit shares for the entry/exit and direction choice probabilities Equations 11 and 12. On movement I assume that, when they start searching in different regions, drivers can choose a direction of motion from the set $\{\text{North}, \text{East}, \text{South}, \text{West}, \varnothing\}$. Direction $\varnothing$ represents the choice of waiting, meaning not moving in any direction, so that $\tau_{\varnothing,ij} \equiv 0$ for every origin-destination pair ij. For $h \neq \varnothing$, I estimate the relocation rates $\tau_{h,ij}$ from the observed cars' GPS trajectories, as the probability that a driver at i moving in direction h enters region j within one minute. I allow these estimates to vary depending on hour and weekday, in order to capture the dynamic variation in traffic speeds.

The remaining parameters to estimate are the time cost $\nu(x)$ for each weekday-hour x, the per-mile transport cost c^d, and drivers' outside option $\bar{u}$. Allowing for a measurement error $\epsilon_i^m(x)$, Equation 26

yields a simple linear equation

$$\sum_j \theta_{ij}^m(x) r_{ij}^m(x) = \delta_i^m(x) + \nu(x)[1 + \sum_j \theta_{ij}^m(x) t_{ij}] + c^d \sum_j \theta_{ij}^m(x) d_{ij} + \epsilon_i^m(x)$$

All variables on the left and right hand sides are observed. In particular, the mean payoffs $\delta_i^m(x)$ can be recovered from observed quantities by exploiting the inversion algorithm described in previous section. The only unknowns are the parameters $\nu(x)$ and c^d, which I estimate by means of a simple linear regression. Finally, drivers' outside option can be recovered from the equilibrium free entry condition $\boldsymbol{u}(\delta) = \bar{u}$, where $\boldsymbol{u}(\delta)$ can be computed by solving drivers' dynamic problem under δ.

5.2.4 Estimates

Figure 10 plots $\nu(x)$ by day of the week and hour of the day. On average, drivers face an opportunity cost of time of about 35$ per hour. This cost is highest at night and on weekends, and lowest during mid-day and on weekdays, which supports an interpretation of it in terms of utility from leisure. Another interpretation is in terms of wages associated to drivers' outside options. One good benchmark is the hourly wage of construction workers in the state of New York, which averaged about 38$ per hour in June 2019 according to official statistics.[33]

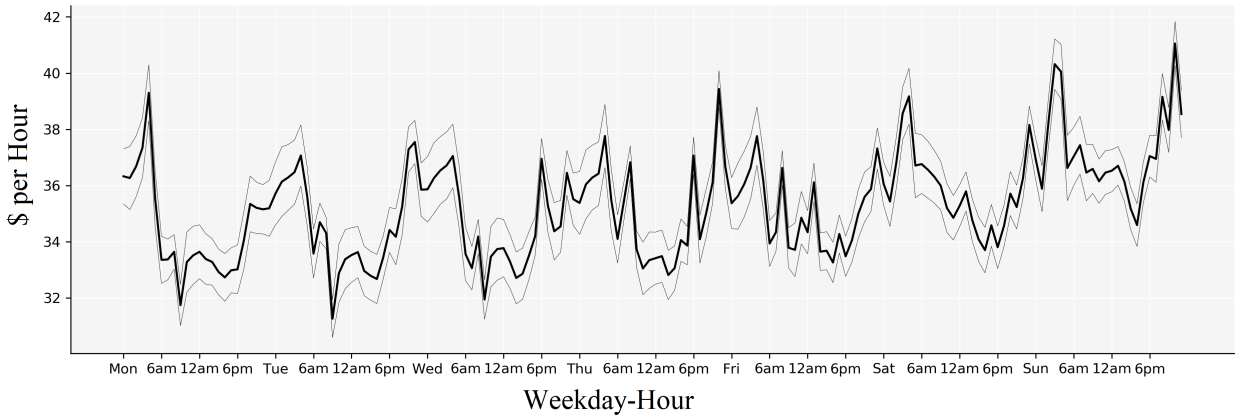

Figure 10: Drivers' value of time as a function of hour and weekday (the shaded lines represent the 95% confidence intervals)

[33] See https://labor.ny.gov/stats/ceshourearn2.asp

I find an estimated driving cost per mile of 0.57\$. As a benchmark, the IRS standard rate for computing deductible costs of operating an automobile for business was 0.58\$ per mile in 2019.[34] The estimated drivers' outside option is 919\$ per year. Recall that this is interpreted as the direct cost for drivers of holding a license. As a comparison, the direct cost of getting a TLC Drivers License ranges between 600\$ and 700\$, while I estimate that the direct costs of maintaining such license are of about 400\$ per year.[35]

5.3 Estimating Platforms' Objective

In my empirical specification I do not assume that platforms fully maximize profits, but instead I estimate an objective function which is a convex combination of profits and customers' welfare. I let λ^m be m's weight on profit maximization, which is to be estimated. Assumptions of this type has been used before in the empirical literature, when profit margins are observed.[36] The reason for estimating such an objective is that, in practice, platforms face incentives to keep prices lower that the full profit maximization level, such as regulatory and media scrutiny. Castillo [2019] argued that such an objective might also capture platforms' incentive to keep prices lower in the short run in order to expand the market in the long run. Moreover notice that, thanks to the observation of prices on both sides, the model primitives were estimated without imposing any assumption on platforms' optimizing behavior. Hence estimating an objective function also allows to model to better fit the observed mark-ups.

Under this objective, platforms set prices according to the equation

$$\begin{aligned} p_{ij}^m(x) =& t_{ij}(\nu(x)+\bar{u}) + c^d d_{ij} + \lambda^m q_{ij}^m(x) \left| \frac{\mathrm{d}\boldsymbol{p}_{ij}^m}{\mathrm{d}q_{ij}^m}(x) \right| \\ &+ W_i^m(x) - (1 - t_{ij}\chi_x) W_j^m(x) - t_{ij} \sum_{x'} \chi_{x,x'} V_j^m(x') \end{aligned} \tag{29}$$

This is the same formula as the one obtained under profit maximization (Equation 22), except for the weights λ^m, $m = a, b$ on the market power distortion, which are to be estimated. I estimate Equation 29 via OLS, finding $\lambda^a = 0.301$ and $\lambda^b = 0.382$. For comparison, Castillo [2019] finds a profit maximization

[34] See https://www.irs.gov/tax-professionals/standard-mileage-rates. I thank Frank Pinter for pointing this out.

[35] For the cost of getting a license, see https://www1.nyc.gov/site/tlc/drivers/get-a-tlc-drivers-license.page. For the yearly costs maintaining a license, a TLC driver license costs 252\$ every three years to renew, and all drivers must bring car in for inspection every other year, paying a TLC vehicle renewal fee of 550\$ and and inspection fee of 75\$.

[36] See Castillo [2019] for instance.

weight of about 0.48 for the monopolist ride-sharing platform in Houston.

6 Counterfactuals

6.1 Merger

In the first counterfactual, I simulate the impact of a merger between the two platforms. I operationalize the merger by assuming that the monopolist platform keeps providing both services a and b to customers, while the driver platforms are integrated in a single one.[37] This allows to abstract from changes in customer welfare due to service variety, focusing instead on the welfare impact due to changes in prices and waiting times.

I begin by an aggregate comparison of the observed equilibrium and the simulated equilibrium under monopoly. The results are shown in Table 4. As reported the first row, the aggregate customer surplus decreases by 2% after the merger. This is the result of a trade-off between a 10% increase in prices and a 10% decrease in waiting times. Both the number of licensed drivers and the number of active drivers (drivers working in a given instant, either matched or idle) drop by 12%. The number of idle drivers drops by 26%, more than twice as much as the number of active drivers, meaning that drivers spend a lower share of their working time waiting for trip requests. This is reflected by an increase in the utilization rate (the fraction of time drivers spend transporting a passenger out of the total time they are active) of about 10%. The annual profits of the monopolist are 29% higher than the total profits observed before the merger, an increase of about 269 M$ per year, which more than compensates for the annual loss in customer welfare. Finally, recall that the average driver payoff associated with holding a driving license is pinned down by the free entry condition, so that the total welfare of the potential drivers in the economy remains constant by assumption. A proxy for drivers' welfare loss can be obtained by looking at the drop in the *gross* drivers' payoff. There are about 10,200 fewer licensed drivers under monopoly, and each license generates a gross payoff of about 900$ per year (drivers' outside option). Hence the decrease in gross drivers' payoffs is about 9 M$ per year.

These results relay an important message: merging the two platforms produces substantial efficiency gains (of about 150 M$ per year), although all gains are captured by the monopolist through increased

[37] In particular, I let the monopolist's profit-maximization weight be equal to the one estimated for the largest platform.

	Change Post-Merger
Annual Customer Surplus	-2% (-117 M$)
Annual Profits	+29% (+269 M$)
Passenger Prices	+10%
Waiting Times	-10%
Licensed Drivers	-12%
Active Drivers	-12%
Idle Drivers	-26%
Driver Utilization Rate	+9.9%
Passenger Trips	-1.4%

Table 4: Merger counterfactual. Changes in market outcomes with respect to their pre-merger values.

profits, leaving users on both sides worse off. These gains are due to the inter-operability of drivers across services that results from the merger. By joining drivers' platforms into a single one, the monopolist can simultaneously increase drivers' matching rates on one side, and decrease passengers' waiting times on the other side.

On aggregate, the efficiency gains resulting from the more efficient use of drivers more than compensate the distortion from increased market power. This crucially relies on several factors, such as market density, passengers' sensitivity to prices and waiting times, and the market concentration prior to the merger. This can be seen from the maps in Figure 12, showing the breakdown of the effects across different regions in the city, and Figure 13, showing correlations explaining the geographical variation. Looking at Figure 12, it can be seen that passengers in peripheral regions gain from the merger. More precisely, Figure 13 shows that the percentage customer welfare loss in different regions is inversely related to the average waiting time before the merger. This is due to the fact that the monopolist draws from a larger pool of drivers with respect to individual platforms, and can exploit this to improve service availability in peripheral regions. Indeed, Figure 13 shows that passengers in regions where supply was scarcer before the merger (higher waiting times) experience a higher percentage decrease in waiting times. The effect on prices, hence the geography of customer welfare loss, also relates to geographical differences in market concentration before the merger. Indeed Figure 13 shows that the average price increase is inversely related to (and well predicted by) the pre-merger market concentration.

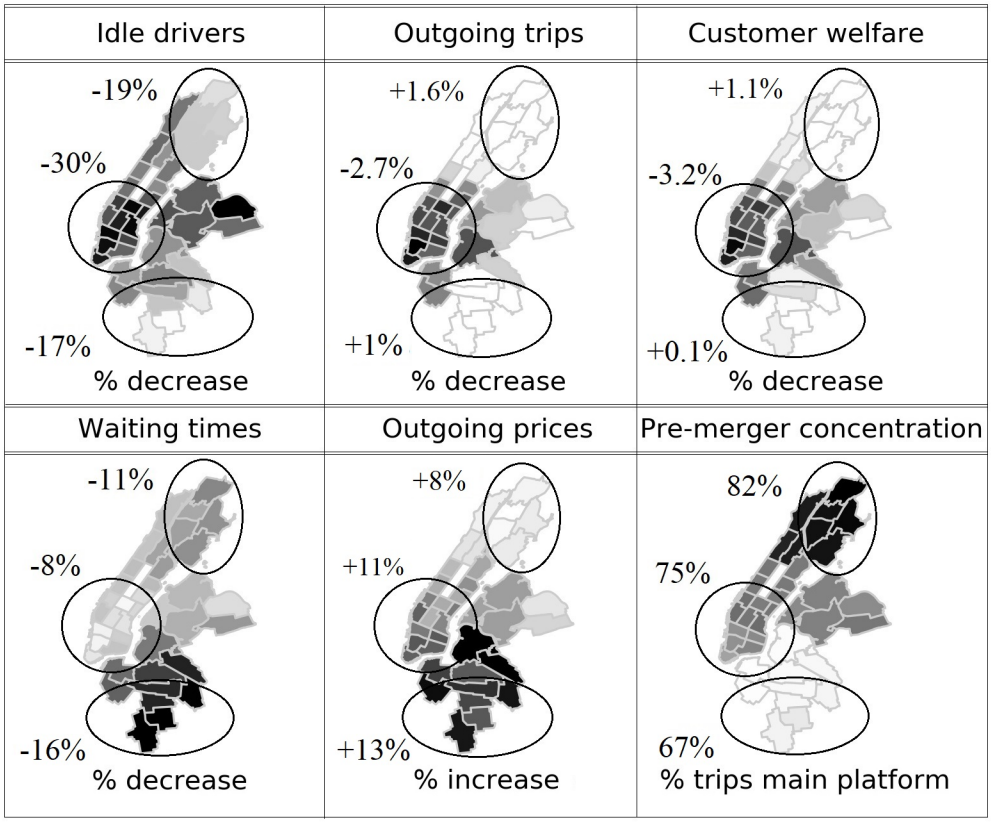

Figure 12: Geography of the effects of the merger. Percentage changes are with respect to pre-merger levels.

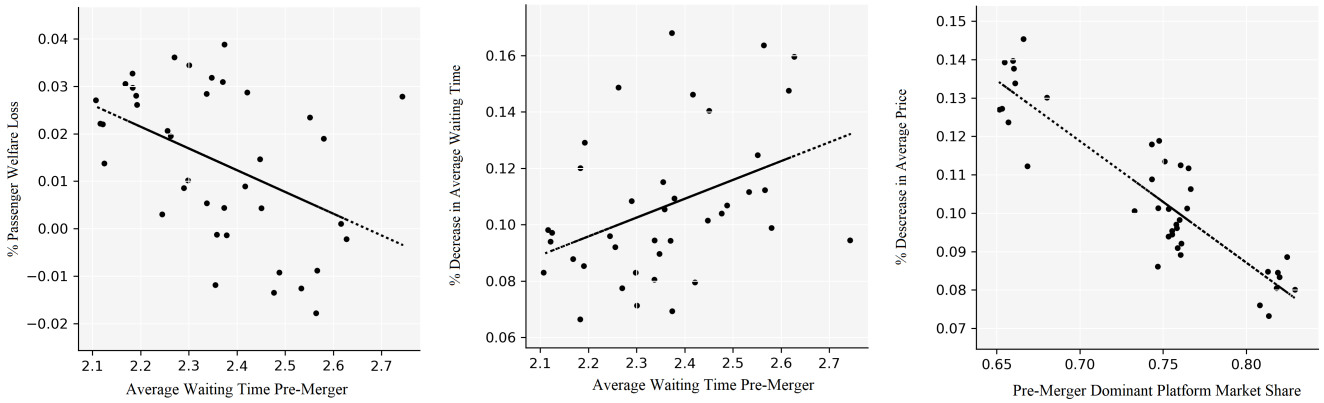

Figure 13: Explaining the geography of the impact of the merger. Average waiting times per-merger can be taken as a proxy for the market density in different regions. Passengers' surplus and waiting times are less affected in less dense regions. Differences in price changes are well-explained by the market concentration prior to the merger.

Finally, there is a significant geographical variation in the decrease in idle drivers, which is higher for the more central and denser regions. Idle vehicles in the CBD decrease by 30%, while the number of trips decrease by only 2.7%. Hence the monopolist provides almost the same number trips as before, but he does so drawing from a significantly lower number of active vehicles. This has a substantial direct impact on traffic emissions. Assuming an average speed of 5 miles per hour, the total distance driven by TNC vehicles in the CBD drops by about 48,000 miles per day. Assuming an average emission of 400 grams of co2 per mile, this translates into a decrease of co2 emissions of bout 19 tons per day in the CBD alone. The same calculation shows that co2 emissions decrease by about 44 tons per day in the whole sample region. This has a significant impact on traffic congestion as well. Schaller [2017] estimates that taxis and TNCs combined comprise 50 percent to as much as 75 percent of all traffic in the CBD, depending on time of day. Taking 40% as a conservative estimate of the average share of TNCs implies that CBD traffic (including all vehicles, not just TNCs) would decline by 5% percent after the merger.

6.2 Congestion Pricing

In recent years, several authors, agencies and media have expressed concerns about the impact of the rise of TNCs on traffic congestion in major US cities. Moreover, it has been shown that TNCs mostly divert passengers from public transit systems, with a consequent lack of financing for transit improvements. In order to fight congestion and provide a financial source for improving public transit, a congestion pricing tax is expected to be implemented in NYC in 2022. The proposed scheme targets fixed costs of access to the CBD, by charging vehicles entering the area. In this section I argue that this type of schemes are unlikely to substantially affect TNCs activity in the CBD. I do so by comparing congestion pricing schemes targeting fixed and variable costs of access. In particular, I simulate the impact of charging drivers 8.3¢ per minute they spend in the CBD (5$ per hour), and compare it with the impact of a tax of 5$ per access to the CBD.[38] Both taxes are charged only between 8 am and 8 pm. As we will see shortly, these tax amounts happen to generate similar revenues in equilibrium, providing a basis for comparison between them. In what follows, I will refer at these schemes as “Per-Minute Tax” and “Entry Tax”, respectively.

[38]The Manhattan meter parking rates, ranging from 4$ to 7.5$ per hour for passenger vehicles, provide a good benchmark.

Under the per-minute tax, drivers are incentivized to reduce the total time they spend inside the CBD, since they are taxed on a per-minute basis regardless on how many times they cross the border. Under the entry tax, they are disincentivized to access the area, but once inside there is no incentive to move out.[39]

Table 5 shows the results. As reported the first row, the two schemes generate a similar tax revenue, of about 60 M$ per year. The pass-through (percentage of the tax revenues paid by passengers through higher prices) is higher for the entry tax than the per-minute tax. Consequently, the entry tax produces a higher passenger welfare loss and a lower reduction in total profits.

The results show that the per-minute tax is significantly more successful at reducing the number of vehicles in the CBD, achieving a 8.5% reduction, compared to a 1.3% reduction under the entry tax. This is obtained primarily by generating a consistent reduction in the number of idle drivers: these drop by 20% under the per-minute tax, compared to 8.5% under the entry-tax.

	Entry Tax	Per-Minute Tax
Annual Tax Revenue	60 M$	55 M$
Annual Passenger Surplus	-1.2% (-69 M$)	-1% (-57 M$)
Annual Profits	-2.6% (-24 M$)	-4% (-38 M$)
Active Drivers in the CBD (8 am - 8pm)	-1.3%	-8.5%
Idle Drivers in the CBD (8am - 8pm)	-3%	-20%
Driver Utilization Rate in the CBD (8am - 8pm)	+0.9%	+6.8%
Licensed Drivers	-1.2%	-2.7%
Waiting Times in the CBD (8am - 8pm)	+0.6%	+5.6%
Trips to the CBD (8am - 8pm)	-2.6%	-2.4%
Trips from the CBD (8am - 8pm)	-0.5%	-2.8%
Average Price for Trips to the CBD (8am - 8pm)	+4.5%	+2.7%
Average Price for Trips from the CBD (8am - 8pm)	+0.7%	+2.9%
Net trip arrivals to the CBD (8am - 8pm)	209 per hour (-42%)	377 per hour (+4.7%)

Table 5: Congestion pricing counterfactuals. Changes in market outcomes with respect to their pre-tax values.

To see why the entry tax is unsuccessful at reducing TNC traffic in the CBD one must look at travel patterns, and in particular to the number of net trip arrivals - defined as the difference between trip

[39] In recent years several cities in Europe have introduced similar schemes. For comparison, non residents are taxed 5€ for accessing Milan's Area C, and 15£ for accessing London's Congestion Charge zone. One slight difference is that these taxes are on a daily basis, while in my simulation drivers are taxed at every access.

arrivals and departures. By increasing the costs of incoming trips, the entry tax reduces net arrivals by 42%. However these remain positive, averaging 209 per hour. This implies that the number of empty vehicles in the CBD remains high even though no idle driver actively relocates inside. Intuitively, car supply in the CBD draws mainly on drivers who remain idle after drop-off, rather than on a net inflow of idle drivers from outer regions. This implies that, *in order to reduce the number of vehicles in the CBD, the entry tax must be strong enough to reverse the natural commuting patterns.* In contrast, a per-minute tax incentivizes a more efficient use of drivers in the CBD by targeting idling time. By forcing idle drivers to relocate outside the CBD right after drop-off, a mild per-minute tax achieves a substantial traffic reduction simply by providing platforms the right incentives to increase waiting times. This situation is shown in detail by the maps in Figure 14, which displays the geographic breakdown of the effects on prices, waiting times and passenger surplus.

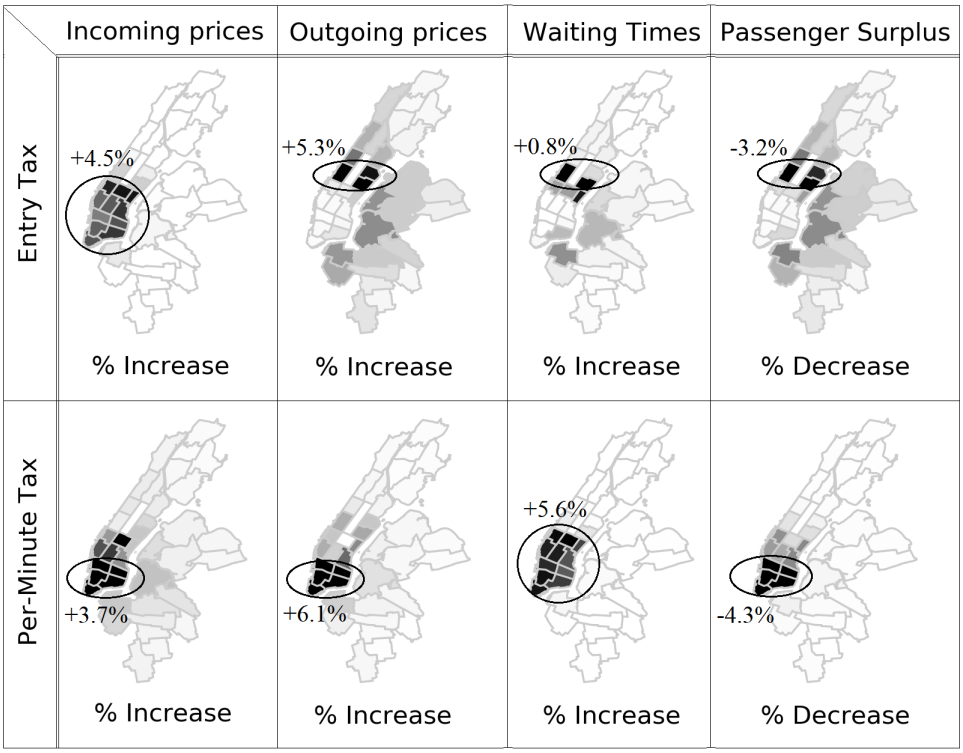

Figure 14: Geography of the effects of congestion pricing. Percentage changes are with respect to pre-tax levels.

From the maps in the first row, it can be seen very clearly how the entry tax reduces the net arrivals in the CBD by raising the prices of incoming trips from outer regions. This produces a loss in passenger surplus in the outer regions, while the passenger surplus in the CBD is not affected. This figure is reversed under the per-minute tax, which has the effect of increasing prices for incoming and outgoing trips symmetrically. Within the CBD, prices increase more for regions in the interior than for regions at the border, and the same is true for waiting times and for the passenger welfare loss.

In conclusion, these results relay a simple message: different congestion pricing schemes will be more or less successful at decreasing TNC traffic in the CBD depending on whether they provide the right incentives to platforms to decrease drivers' idling time. These incentives are provided by schemes targeting variable costs of access to the CBD. Schemes targeting fixed costs, such as the one which which will go into place in 2022, are likely to not produce a significant change in TNC activity in the CBD, but instead they are likely to affect ridership and customer welfare mostly in outer regions.

6.3 Entry Restrictions

In the last counterfactual, I study the impact of measures restricting driver supply. I do so by simulating the market equilibrium under a tax on driving licenses of 10,000$ per year.[40] This is operationalized through the free entry condition, by imposing that in equilibrium the annualized optimal expected average payoff of licensed drivers must be equal to the license tax . The aggregate results are presented in Table 6. The first thing to note is that the number of licensed drivers drops by 84%, from about 85,000 to 13,757 licensed drivers. This provides a basis for comparison with the current medallion system in place for yellow taxi cabs (the total number of medallions is currently 13,587 in NYC). From the first three rows we can see that, according to the simulation, such tax would produce a total welfare loss of about 200M$ per year, trading off a sizable tax revenue with large reductions in customers' surplus and platforms' profits. Under the tax, however, drivers' utilization rate increases by 3.3%, producing a sizable reduction in the number of active vehicles in the city. This is achieved primarily through a large decrease in empty vehicles, but also through a sizable decrease in customer trips. Finally, note that the decrease in the average number of active vehicles is one order of magnitude lower than the decrease in the number of

[40] For comparison, the current lease cap rate for yellow taxi cab medallions is about 600$ per week, corresponding to about 30000$ per year. See https://www1.nyc.gov/assets/tlc/downloads/pdf/rule_book_current_chapter_58.pdf

licenses, meaning that the latter are used more intensively under the new tax, which is intuitive.

	Change Post-Tax
Annual Tax Revenue	124 M$
Annual Customer Surplus	-3.9% (-233 M$)
Annual Profits	-9.4% (-88 M$)
Passenger Prices	+6.1%
Waiting Times	+3.7%
Licensed Drivers	-84%
Active Drivers	-6.7%
Idle Drivers	-12%
Driver Utilization Rate	+3.3%
Passenger Trips	-3.8%

Table 6: License Tax. Changes in market outcomes with respect to their pre-tax values.

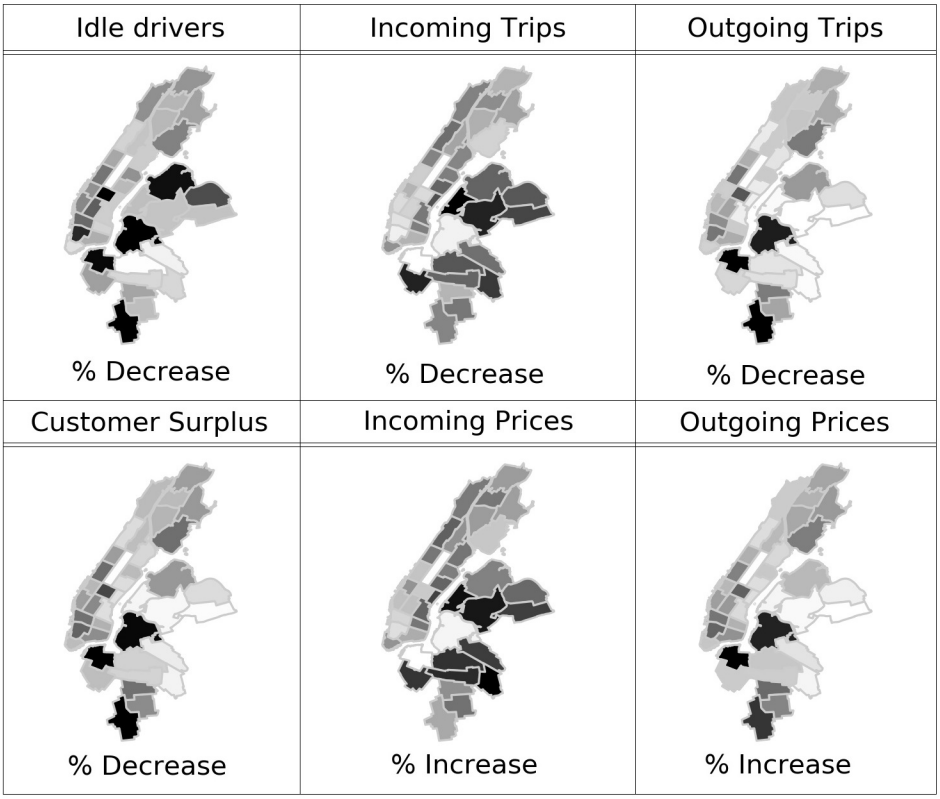

Figure 15: Geography of the effects of a license tax. Percentage changes with respect to pre-tax levels.

Figure 15 shows the effects disaggregated by region. One concern which is often raised in relation to policies restricting supply, is that these would disproportionately affect customers in peripheral regions. Indeed, one main benefit associated with ride-sharing platforms is that these have been able to expand the access of taxi services to regions in the periphery that were previously excluded. One argument supporting this concern is that markets of this type feature "economies of density", meaning that the quality of the match increases with the market size, hence this type of policies would produce a concentration of the activities in the densest central regions. Figure 15 shows that this is not the case under this simulation. Although there is a considerable degree of dispersion in how the tax impacts different quantities, there is no clear pattern across different boroughs.

To understand why this is the case, notice first that the maps in Figure 15 are specular images one of another. This is shown by the scatter plots in Figure 16. Figures 16.a and 16.b show that regions where outgoing prices increase more are those where incoming prices increase less, and the same is true for incoming and outgoing trips. That is, regions can be partitioned in two sets: regions that are penalized as trip origins, and regions that are penalized as trip destinations. Figure 16.c shows that regions penalized as origins are also those where car supply decreases the most - which mechanically translates into a higher increase in waiting times as well, hence a higher customer welfare loss.

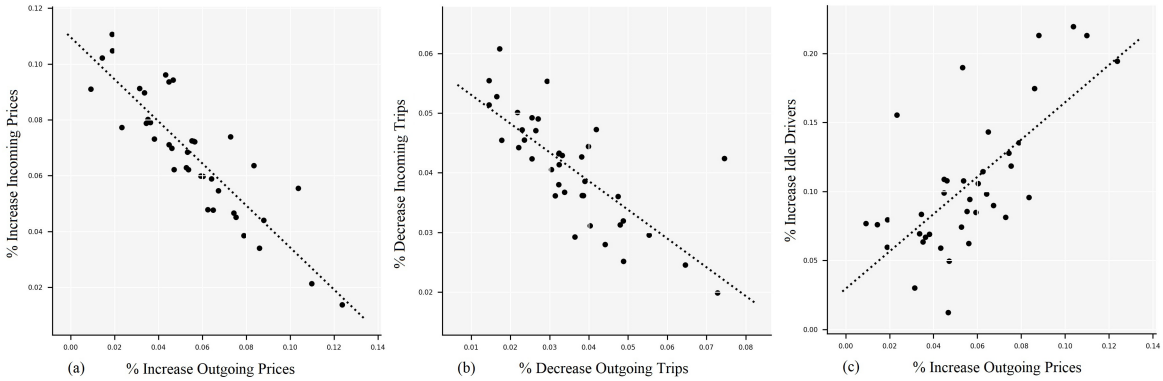

Figure 16: Explaining the geography of the effects of a license tax. The maps in Figure 15 are the specular images one of another, meaning that certain regions are penalized as origins, while other regions are penalized as destinations.

The cause of this lies in the network of commuting patterns, and the driver movements that these

patterns induce. The most affected regions are the "trip attractors", meaning those regions that attract more trips that they originate. This is because supply in these regions largely draws on drivers who remain idle after drop-off. Intuitively, before the tax, these regions were able to capture those drivers who dropped a passenger and did not receive a new request. After the tax the number of trips generally drops, meaning that these regions are not able to capture a large number of net arrivals anymore. Symmetrically, regions generating a large number of net departures are the least affected by the tax. This is because these regions draw mainly on a net inflow of idle drivers, a source of supply which is less affected by the price increase. This situation is shown in Figure 17. Figure 17.a shows that the license tax does not affect the direction of drivers' net relocation flows. This is intuitive, since drivers' movements must balance passenger's commuting patterns, whose direction is not affected by the price increase. Figure 17.b shows that regions who draw their car supply from a net inflow of idle drivers (as opposed to net trip arrivals) are the least affected by the tax.

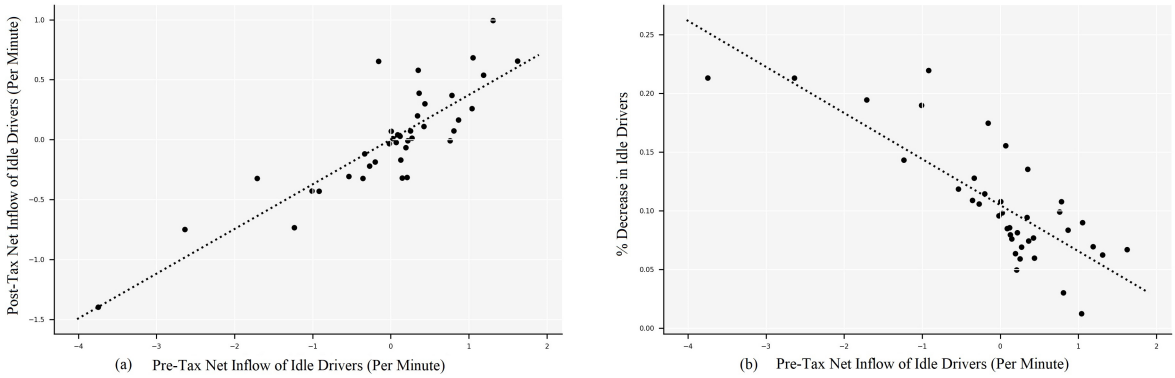

Figure 17: Explaining the geography of the effects of a license tax. Regions that draw their supply from relocating drivers (hence are "trip generators") are less affected. In particular, they keep receiving relocating drivers (panel a) and experience a lower percentage reduction of car supply (panel b).

In summary, these pictures reach to a simple conclusion: in the current context, it is not clear that policy measures restricting car supply would affect mostly the periphery, since platforms are able to coordinate a smaller pool of licensed drivers so to avoid a concentration of activities in central regions. To understand the geographical variation in the impact of such policies market density is not crucial, but instead it is important to understand how car supply relates to travel patterns.

7 Conclusion

This paper studies competition in the app-based transportation industry. I present a model of competing platforms in transport equilibrium, characterizing analytically the profit-maximizing allocations and prices. I construct a novel dataset covering two months of high-frequency data on the operations of the two main platforms in New York City, and estimate the model exploiting a novel methodology and a policy change. I then use the model to study different regulations, and a merger between platforms. My counterfactuals suggest that, realizing economies of density, a merger would improve efficiency and reduce traffic volumes. They also indicate the importance of targeting variable instead of fixed costs of access in order to reduce traffic volumes in central areas, and they suggest that entry restrictions do not disproportionately affect the peripheries, but can generate a large efficiency loss. Finally, they indicate the importance of market density and the network of travel patterns to understand the geographic variation of the outcomes of regulation.

References

Joshua D Angrist, Sydnee Caldwell, and Jonathan V Hall. Uber vs. taxi: A driver's eye view. *National Bureau of Economic Research*, 2017.

Mark Armstrong. Competition in two-sided markets. *The RAND Journal of Economics*, 37(3):668–691, 2006.

Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

Steven Berry, Amit Gandhi, and Philip Haile. Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111, 2013.

Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

Giulia Brancaccio, Myrto Kalouptsidi, and Theodore Papageorgiou. Geography, transportation, and endogenous trade costs. *Econometrica*, 88(2):657–691, 2020a.

Giulia Brancaccio, Myrto Kalouptsidi, Theodore Papageorgiou, and Nicola Rosaia. Search frictions and efficiency in decentralized transport markets. 2020b.

Nicholas Buchholz. Spatial equilibrium, search frictions and dynamic efficiency in the taxi industry. *mimeo, Princeton University*, 2020.

Nicholas Buchholz, Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz. The value of time: Evidence from auctioned cab rides. *National Bureau of Economic Research*, 2020.

Guangyu Cao, Ginger Zhe Jin, Xi Weng, and Li-An Zhou. Market expanding or market stealing? competition with network effects in bikesharing. *Working paper, National Bureau of Economic Research*, 2018.

Joe Castiglione, Drew Cooper, Bhargava Sana, Dan Tischler, Tilly Chang, Gregory D Erhardt, Sneha Roy, Mei Chen, and Alex Mucci. Tncs & congestion. 2018.

Juan Camilo Castillo. Who benefits from surge pricing? *mimeo, Stanford University*, 2019.

Ambarish Chandra and Allan Collard-Wexler. Mergers in two-sided markets: An application to the canadian newspaper industry. *Journal of Economics & Management Strategy*, 18(4):1045–1070, 2009.

Le Chen, Alan Mislove, and Christo Wilson. Peeking beneath the hood of uber. In *Proceedings of the 2015 internet measurement conference*, pages 495–508, 2015.

M Keith Chen, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen. The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 127(6):2735–2794, 2019.

Khai Xiang Chiong, Alfred Galichon, and Matt Shum. Duality in dynamic discrete-choice models. *Quantitative Economics*, 7(1):83–115, 2016.

Matthew T Clements and Hiroshi Ohashi. Indirect network effects and the product cycle: video games in the us, 1994–2002. *The Journal of Industrial Economics*, 53(4):515–542, 2005.

Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. *National Bureau of Economic Research*, 2016a.

Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. 2016b.

Cody Cook, Rebecca Diamond, Jonathan Hall, John A List, and Paul Oyer. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *National Bureau of Economic Research*, 2018.

Kenneth S. Corts and Mara Lederman. Software exclusivity and the scope of indirect network effects in the us home video game market. *international Journal of industrial Organization*, 27(2):121–136, 2009.

Guillaume R. Frechette, Alessandro Lizzeri, and Tobias Salz. Frictions in a competitive, regulated market evidence from taxis. *American Economic Review*, 109(8):2954–2992, 2019.

Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

Neil Gandal, Michael Kende, and Rafael Rob. The dynamics of technological adoption in hardware/software systems: The case of compact disc players. *The RAND Journal of Economics*, pages 43–61, 2000.

Soheil Ghili and Vineet Kumar. Spatial distribution of supply and the role of market thickness: Theory and evidence from ride sharing. *mimeo, Yale University*, 2020.

Michael Graehler, Richard Alexander Mucci, and Gregory D Erhardt. Understanding the recent transit ridership decline in major us cities: service cuts or emerging modes. In *Transportation Research Board 98th Annual Meeting, Washington, DC, January*, 2019.

Jonathan V Hall, John J Horton, and Daniel T Knoepfle. Pricing efficiently in designed markets: The case of ride-sharing. *Available john-joseph-horton. com*, 2019.

Arthur J. Hosios. On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, 57(2):279–298, 1990.

Przemysław Jeziorski. Effects of mergers in two-sided markets: The us radio industry. *American Economic Journal: Microeconomics*, 6(4):35–73, 2014.

Shan Jiang, Le Chen, Alan Mislove, and Christo Wilson. On ridesharing competition and accessibility: Evidence from uber, lyft, and taxi. In *Proceedings of the 2018 World Wide Web Conference*, pages 863–872, 2018.

Ulrich Kaiser and Julian Wright. Price structure in two-sided markets: Evidence from the magazine industry. *International Journal of Industrial Organization*, 24(1):1–28, 2006.

Gabriel Kreindler. Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. *Working paper, Harvard University*, 2020.

Ricardo Lagos. An alternative approach to search frictions. *Journal of Political Economy*, 108(5):851–873, 2000.

Robin S Lee. Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7):2960–3000, 2013.

Tracy Xiao Liu, Zhixi Wan, and Chenyu Yang. The efficiency of a dynamic decentralized two-sided matching market. *mimeo, University of Rochester*, 2019.

Daniel McFadden. Modeling the choice of residential location. *Transportation Research Record*, (673), 1978.

Alejandro Molnar and Daniel Mangrum. The marginal congestion of a taxi in new york city. *Working paper*, 2018.

Ariel Pakes. Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica*, (54):755–784, 1986.

Ariel Pakes, Michael Ostrovsky, and Steven T. Berry. Simple estimators for the parameters of discrete dynamics games. *The RAND Journal of Economics*, 38(2):373–399, 2007.

James A. Parrott and Michael Reich. An earnings standard for new york city's app-based drivers: Economic analysis and policy assessment. Technical report, 2018.

Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029, 2003.

Nicola Rosaia. The patient harold zurcher and his dual: Estimating undiscounted markov decision processes. *Working paper*, 2020.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5): 999–1033, 1987.

John Rust. Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143, 1994.

Marc Rysman. Competition between networks: A study of the market for yellow pages. *The Review of Economic Studies*, 71(2):483–512, 2004.

Bruce Schaller. Empty seats, full streets. fixing manhattan's traffic problem. Technical report, 2017.

SFCTA. Today: A profile of san francisco transportation network company activity, 2017.

Matthew H. Shapiro. Density of demand and the benefit of uber. *mimeo, Singapore Management University*, 2018.

A Michael Spence. Monopoly, quality, and regulation. *The Bell Journal of Economics*, pages 417–429, 1975.

E. Glen Weyl. A price theory of multi-sided platforms. *American Economic Review*, 100(4):1642–72, 2010.

A Proofs

A.1 Preliminaries

A.1.1 Williams-Daly-Zachary

This section introduces results from optimal transport theory applied to discrete choice, which will be later used in the proofs, drawing from Galichon [2018]. The result of this section is a slight re-statement of the well-known Williams-Daly-Zachary Theorem, an analogous to Roy's identity in discrete-choice models which has been used, for instance, by McFadden [1978], Rust [1994, Theorem 3.1] and Chiong et al. [2016].

Abstracting from dynamic considerations for now, consider a driver facing the entry/exit decision problem at x, i, and let $v = (v^m)_{m \in M \cup \{out\}}$ be a vector of continuation values associated to different choices. The inclusive value of the entry/exit choice under v is given by

$$V(v) = \mathbb{E} \max_{m \in M \cup \{out\}} (v^m + \frac{\eta^m}{\chi_x})$$

and the optimal vector of entry/exit choice probabilities $\mathrm{p} = (\mathrm{p}^m)_{m \in M \cup \{out\}}$ must satisfy

$$\mathrm{p}^m = \mathbb{P}[v^m + \frac{\eta^m}{\chi_x} = \max_{m' \in M \cup \{out\}} (v^{m'} + \frac{\eta^{m'}}{\chi_x})] \qquad \forall m \in M \cup \{out\} \tag{30}$$

For any given vector $a \in \mathbb{R}_+^{M \cup \{out\}}$, define the choice probabilities $\mathrm{p}(a)$ associated with it by

$$\mathrm{p}^m(a) = \frac{a^m}{\sum_{m' \in M \cup \{out\}} a^{m'}} \qquad \forall m \in M \cup \{out\}$$

and say that V, v, a *is optimal for the entry/exit choice at* x, i if $\mathrm{p}(a)$ satisfies 30 and $V = V(v)$. Intuitively the sum $\sum_{m \in M \cup \{out\}} a^m$ can be seen as representing the total number of drivers facing the entry/exit choice at x, i, which is taken as exogenous for now, while later it will be determined endogenously by drivers' movements. $\mathrm{p}(a)$ represents the probabilities according to which these drivers choose to start working on different platforms m (or become inactive if $m = out$). v and V represent the vector of continuation values associated to different choices and the inclusive value of the entry/exit choice, respectively. Optimality requires V, v, a to be consistent with a static discrete choice setup. Next result characterizes

optimality if V, v, a in terms of convex conjugacy. For every vector of probabilities p define the function:

$$g^{\mathrm{e}}(\mathrm{p}) \equiv \max_{v \in \mathbb{R}^{M \cup \{out\}}} [\sum_{m \in M \cup \{out\}} \mathrm{p}^m v^m - \mathbb{E} \max_{m' \in M \cup \{out\}} (v^{m'} + \eta^{m'})]$$

and for every $a \in \mathbb{R}_+^{M \cup \{out\}}$ define the function

$$G^{\mathrm{e}}(a) = \sum_{m \in M \cup \{out\}} a^m g^{\mathrm{e}}(\mathrm{p}(a))$$

Proposition 2. *G^e is strictly convex, differentiable and such that*

$$\sum_m a^m \frac{dG^e(a)}{da^m} = G^e(a)$$

Moreover, V, v, a is optimal for the entry/exit choice at x, i if and only if

$$\frac{dG^e(a)}{da^m} = \chi_x(v^m - V) \qquad \forall m \in M \cup \{out\}$$

Proof. That G^{e} is strictly convex and differentiable follows from the assumption that η has continuous density and full support (See Galichon [2018, Proposition D.14]). For the second part, denote by $\frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p})}{\mathrm{dp}} \subseteq \mathbb{R}^{M \cup \{out\}}$ the super-gradient of g^{e} at p.[41] From Galichon [2018, Propositions 9.8 and D.13] it follows that V, v, a is optimal if and only if

$$\chi_x v \in \frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p}(a))}{\mathrm{dp}} \tag{31}$$

and that the latter is equivalent to

$$g^{\mathrm{e}}(\mathrm{p}(a)) + \chi_x V = \chi_x \sum_{m \in M \cup \{out\}} \mathrm{p}^m(a) v^m \tag{32}$$

[41] Formally function g^{e} is not differentiable, but it can be shown that it is such that, for every $x, y \in \frac{\mathrm{d}g^{\mathrm{e}}}{\mathrm{dp}}$, there exists a scalar $k \in \mathbb{R}$ such that $x = y + k$. Therefore for every $x, y \in \frac{\mathrm{d}g^{\mathrm{e}}}{\mathrm{dp}}$ we have

$$-\sum_{m'} \mathrm{p}^{m'}(a) x^{m'} + x^m = -\sum_{m'} \mathrm{p}^{m'}(a) y^{m'} + y^m$$

which justifies notation in CIT.

Notice that, in turn, this is equivalent to

$$\chi_x(v^m - V) = g^{\mathrm{e}}(\mathrm{p}(a)) - \sum_{m'} \mathrm{p}^{m'}(a) \frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m'}} + \frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m}} \tag{33}$$

Indeed, note that if 31 and 32 hold then

$$\begin{aligned}
&g^{\mathrm{entry}}(\mathrm{p}(a)) - \sum_{m'} \mathrm{p}^{m'}(a) \frac{\mathrm{d}g^{\mathrm{entry}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m'}} + \frac{\mathrm{d}g^{\mathrm{entry}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m}} \\
&= g^{\mathrm{entry}}(\mathrm{p}(a)) - \chi_x \sum_{m'} \mathrm{p}^{m'}(a) v^{m'} + \chi_x v^m \\
&= g^{\mathrm{entry}}(\mathrm{p}(a)) - [g^{\mathrm{entry}}(\mathrm{p}(a)) + \chi_x V] + \chi_x v^m \\
&= \chi_x(v^m - V)
\end{aligned}$$

and conversely, if 33 holds then

$$\begin{aligned}
&\chi_x[\sum_m \mathrm{p}^m(a) v^m - V] \\
&= \sum_m \mathrm{p}^m(a)[g^{\mathrm{entry}}(\mathrm{p}(a)) - \sum_{m'} \mathrm{p}^{m'}(a) \frac{\mathrm{d}g^{\mathrm{entry}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m'}} + \frac{\mathrm{d}g^{\mathrm{entry}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m}}] \\
&= g^{\mathrm{entry}}(\mathrm{p}(a))
\end{aligned}$$

Hence the result follows by noting that

$$\frac{\mathrm{d}G^{\mathrm{e}}(a)}{\mathrm{d}a^m} = g^{\mathrm{e}}(\mathrm{p}(a)) - \sum_{m'} \mathrm{p}^{m'}(a) \frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m'}} + \frac{\mathrm{d}g^{\mathrm{e}}(\mathrm{p}(a))}{\mathrm{d}\mathrm{p}^{m}}.$$

□

For every x, i, m, optimality of drivers' direction choices at x, i, m can be analogously characterized. Formally, for every vector of direction choice probabilities $\mathrm{p} = (\mathrm{p}^h)_{h \in H}$ define the convex conjugate

$$g^{\mathrm{d}}(\mathrm{p}) \equiv \max_{v \in \mathbb{R}^H} [\sum_{h \in H} \mathrm{p}^h v^h - \mathbb{E} \max_{h' \in H} (v^{h'} + \epsilon^{h'})]$$

and for every vector $s \in \mathbb{R}_+^H$ define the function

$$G^{\mathrm{d}}(s) = \sum_{h \in H} s^h g^{\mathrm{d}}(\mathrm{p}(s))$$

where $\mathrm{p}^h(s) \equiv s^h / \sum_{h'} s^{h'}$ for each $h \in H$. Say that V, v, s *is optimal for the direction choice* at $x, i, , m$ if $\mathrm{p}(s)$ satisfies

$$\mathrm{p}^h(s) = \mathbb{P}[v^h + \frac{\epsilon^h}{\theta_i^m(x) + \lambda + \chi_x} = \max_{h' \in H}(v^{h'} + \frac{\epsilon^{h'}}{\theta_i^m(x) + \lambda + \chi_x})]$$

and V is given by

$$V = \mathbb{E} \max_{h \in H}[v^h + \frac{\epsilon^h}{\theta_i^m(x) + \lambda + \chi_x}]$$

Similarly to the case of the entry/exit choice, the sum $\sum_H s^h$ can be seen as representing the total number of drivers searching at x, i, m, which is taken as exogenous for now, while later it will be determined endogenously by drivers' movements, and $\mathrm{p}(s)$ represents the probabilities according to which these drivers are headed to different directions. Optimality requires V, v, a to be consistent with a static discrete choice setup. Proceeding as in the proof of previous proposition one can show the following.

Proposition 3. *G^d is strictly convex, differentiable and such that*

$$\sum_h s^h \frac{dG^e(h)}{ds^h} = G^e(s).$$

Moreover, V, v, s is optimal for the direction choice at x, i, m if and only if

$$\frac{dG^d(s)}{ds^h} = [\theta_i^m(x) + \lambda + \chi_x](v^h - V) \qquad \forall h \in H$$

A.1.2 Drivers' Optimal Dual Pairs

Consider the long-run distribution of N drivers across activities in the network associated with σ, θ. This distribution is described by: i) the request rates q, capturing the frequency at which drivers start transporting passengers on different routes; ii) the long-rung distribution of idle drivers $\bar{s}$ and potential

entrants $\bar{e}$. This distribution must be such that $\bar{e} = \bar{\boldsymbol{e}}(\sigma, \theta, N)$, $\bar{s} = \bar{\boldsymbol{s}}(\sigma, \theta, N)$, and for every x, ij, m

$$q_{ij}^m(x) = \bar{s}_i^m(x)\theta_{ij}^m(x)$$

Aggregating Constraints 13 and 14 across N drivers, this is the case if and only if $q, \bar{s}, \bar{e}, \sigma$ satisfies the following constraints for every x, ij, m:

$$\begin{aligned}
&\mu(x)q_{ij}^m(x) = \mu(x)\bar{s}_i^m(x)\theta_{ij}^m(x) \qquad (34)\\
&\mu(x)\bar{e}_i(x)\chi_x = \sum_{x'}\mu(x')[\sum_m \bar{s}_i^m(x') + \sum_{m,j} q_{ji}^m(x')t_{ji} + \bar{e}_i(x')]\chi_{x',x}\sigma_i^{out}(x)\\
&\mu(x)\sum_{ij} q_{ij}^m(x) + \mu(x)\bar{s}_i^m(x)[\sum_{h,j}\sigma_i^{h|m}(x)\tau_{h,ij} + \chi_x] = \mu(x)\sum_j \bar{s}_j^m(x)\sum_h \sigma_j^{h|m}(x)\tau_{h,ji}\\
&\quad + \mu(x)\sum_j q_{ji}^m(x)(1-\chi_x t_{ji}) + \sum_{x'}\mu(x')[\sum_m \bar{s}_i^m(x') + \sum_{m,j} q_{ji}^m(x')t_{ji} + \bar{e}_i(x')]\chi_{x',x}\sigma_i^m(x)\\
&\sum_{m,i}\bar{s}_i^m(x) + \sum_{m,ij} q_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x) = N
\end{aligned}$$

First, I show that the last set of constraints is equivalent to

$$\sum_x \mu(x)[\sum_{m,i}\bar{s}_i^m(x) + \sum_{m,ij} q_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x)] = N$$

To see this, note that summing the second and third set of constraints across i, m implies that for every x

$$\begin{aligned}
&\mu(x)[\sum_{m,i}\bar{s}_i^m(x) + \sum_{m,ij} q_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x)]\chi_x\\
&= \sum_{x'}\mu(x')[\sum_{m,i}\bar{s}_i^m(x') + \sum_{m,ij} q_{ij}^m(x')t_{ij} + \sum_i \bar{e}_i(x')]\chi_{x',x}
\end{aligned}$$

Hence defining

$$\tilde{\mu}(x) = \frac{\mu(x)[\sum_{m,i}\bar{s}_i^m(x) + \sum_{ij} q_{m,ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x)]}{\sum_{x'}\mu(x')[\sum_{m,i}\bar{s}_i^m(x') + \sum_{m,ij} q_{ij}^m(x')t_{ij} + \sum_i \bar{e}_i(x')]}$$

We have $\sum_x \tilde{\mu}(x) = 1$ and $\tilde{\mu}(x) = \sum_{x'} \tilde{\mu}(x')\chi_{x',x}$, hence $\tilde{\mu}$ is a stationary distribution for χ. This implies that

$$\sum_{m,i} \bar{s}_i^m(x) + \sum_{m,ij} q_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x) = \sum_{m,i} \bar{s}_i^m(x') + \sum_{m,ij} q_{ij}^m(x')t_{ij} + \sum_i \bar{e}_i(x')$$

for every x, x'm because otherwise we would have $\tilde{\mu} \neq \mu$, contradicting the fact that χ has a unique stationary distribution.

Second, notice that, for any given $q, \bar{s}, \bar{e}, \sigma$, we can define:

- The long-run entry/exit rates at every x, i

$$\alpha_i^m(x) = \sum_{x'} \mu(x')[\sum_m \bar{s}_i^m(x') + \sum_{m,j} q_{ji}^m(x')t_{ji} + \bar{e}_i(x')]\chi_{x',x}\sigma_i^m(x) \tag{35}$$

  Intuitively, $\alpha_i^m(x)$ is the rate at which drivers choose to start working for m in state x (or to become inactive, if $m = out$), which is the product of the rate at which they face the entry/exit decision times the probability according to which they choose m.

- The expected number $s_{i,h}^m(x)$ of drivers searching at x, i, m headed towards h, in the long-run

$$s_{i,h}^m(x) = \bar{s}_i^m(x)\sigma_i^{h|m}(x) \tag{36}$$

For what follows, it is convenient to work with α, s instead of $\bar{s}, \sigma$. Substituting 35 and 36 into the stationarity constraints in 34, feasibility requires $q, s, \bar{e}, \alpha$ to satisfy the following constraints for every

x, ij, m

$$\mu(x)q_{ij}^m(x) = \mu(x)\sum_h s_{i,h}^m(x)\theta_{ij}^m(x) \tag{37}$$

$$\mu(x)\bar{e}_i(x)\chi_x = \alpha_i^{out}(x) \tag{38}$$

$$\mu(x)[\sum_j q_{ij}^m(x) + \sum_h s_{i,h}^m(x)(\sum_j \tau_{h,ij} + \chi_x)] \tag{39}$$

$$= \mu(x)\sum_{j,h} s_{j,h}^m(x)\tau_{h,ji} + \mu(x)\sum_j q_{ji}^m(x)(1 - \chi_x t_{ji}) + \alpha_i^m(x)$$

$$\sum_m \alpha_i^m(x) = \sum_{x'} \mu(x')[\sum_{m,h} s_{i,h}^m(x') + \sum_{m,j} q_{ji}^m(x')t_{ji} + \bar{e}_i(x')]\chi_{x',x} \tag{40}$$

$$\sum_x \mu(x)[\sum_{m,i,h} s_{i,h}^m(x) + \sum_{m,ij} q_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x)] = N \tag{41}$$

To be consistent with drivers' optimal response to r, θ, $q, s, \bar{e}, \alpha$ must be such that the associated choice probabilities σ are optimal, where the latter can be defined from

$$\sigma_i^m(x) = \frac{\alpha_i^m(x)}{\sum_{m' \in M \cup \{out\}} \alpha_i^{m'}(x)} \qquad \text{and} \qquad \sigma_i^{h|m}(x) = \frac{s_{i,h}^m(x)}{\sum_{h' \in H} s_{i,h'}^m(x)} \tag{42}$$

In particular, there must exist a system of drivers' value functions W, V, U, J and an optimal average payoff flow u satisfying the Bellman Equations 6-8 such that σ is optimal under W, U (i.e. it satisfies Equations 11 and 12). For what follows, it is convenient to slightly re-write the Bellman equations as follows. For every x, ij, m define drivers' dynamic surplus of receiving a trip request by

$$\Delta_{ij}^m(x) = r_{ij}^m(x) - c^d d_{ij} + J_{ij}^m(x) - W_i^m(x)$$

Intuitively, after receiving a trip request the driver gets the lump-sum payoff $r_{ij}^m(x) - c^d d_{ij}$ and the continuation value $J_{ij}^m(x)$, but gives up the opportunity of searching at x, i, m, whose continuation value if given by $W_i^m(x)$. Given this, Equations 6-8 can be re-written as

$$W_i^{out}(x) = -\frac{u}{\chi_x} + \sum_{x'} \frac{\chi_{x,x'}}{\chi_x} V_i(x') \tag{43}$$

$$[\lambda + \theta_i^m(x) + \chi_x][U_{i,h}^m(x) - W_i^m(x)] = -\nu(x) - u + \sum_j \theta_{ij}^m(x)\Delta_{ij}^m(x) \tag{44}$$
$$+ \sum_{x'} \chi_{x,x'}[V_i(x') - W_i^m(x)] + \sum_j \tau_{h,ij}[W_j^m(x) - W_i^m(x)]$$

$$\Delta_{ij}^m(x) = r_{ij}^m(x) - c^d d_{ij} - t_{ij}(\nu(x) + u) \tag{45}$$
$$+ t_{ij} \sum_{x'} \chi_{x,x'} V_j(x') + (1 - t_{ij}\chi_x) W_j^m(x) - W_i^m(x)$$

Moreover, using the results of previous section, optimality of σ and the definitions of the inclusive values in Equations 9 and 10 are equivalent to the following

$$\frac{\mathrm{d}G^{\mathrm{d}}(s_i^m(x))}{\mathrm{d}s_{i,h}^m(x)} = [\theta_i^m(x) + \lambda + \chi_x][U_{i,h}^m(x) - W_i^m(x)] \quad \forall m \in M, h \in H \tag{46}$$

$$\frac{\mathrm{d}G^{\mathrm{e}}(\alpha(x))}{\mathrm{d}\alpha_i^m(x)} = \chi_x[W_i^m(x) - V_i(x)] \quad \forall m \in M \cup \{out\} \tag{47}$$

This yields the following definition.

Definition 8. $(q, s, \bar{e}, \alpha), (W, V, \Delta, u)$ *is an optimal dual pair under* r, θ, N if:

1. $q, s, \bar{e}, \alpha$ satisfies the stationarity constraints 37-41
2. Defining U from W, V, Δ, u as in Equation 44, W, V, Δ, u, U satisfy the Equations 43-47

In words, $(q, s, \bar{e}, \alpha), (W, V, \Delta, u)$ is an optimal dual pair if it fully describes the optimal long-run distribution of drivers under r, θ, and the associated drivers' value functions. Next Section provides a straightforward characterization of such pairs in terms of a regularized linear program.

A.1.3 Characterization

Define drivers' welfare at $q, s, \bar{e}, \alpha$ by

$$f(r,q,s,\bar{e},\alpha) = \sum_x \mu(x)\{\sum_{m,ij} q_{ij}^m(x)[r_{ij}^m(x) - c^d d_{ij} - t_{ij}\nu(x)] - \sum_{m,i,h} s_{i,h}^m(x)\nu(x)\}$$
$$- \sum_{x,i,m} \mu(x) G^{\mathrm{d}}(s_i^m(x)) - \sum_{x,i} \frac{1}{\chi_x} G^{\mathrm{e}}(\alpha_i(x))$$

Intuitively, conditional on state x: i) $q_{ij}^m(x)$ drivers get matched on ij, m each instant on average, receiving $r_{ij}^m(x) - c^d d_{ij}$ lump-sum and paying the time cost $\nu(x)$ for t_{ij} time periods; ii) $\sum_h s_{i,h}^m(x)$ drivers are searching at i, m, headed towards h, paying the time cost flow $\nu(x)$; and iii) The sum on the second line captures drivers' utility flow from the idiosyncratic preference shocks. Consider the problem of maximizing this welfare subject to the stationarity Constraints 37-41:

$$\max_{q,s,\bar{e},\alpha} f(r,q,s,\bar{e},\alpha) \tag{48}$$
$$\text{s.t.} 37-41$$

Theorem 3. *$(q, s, \bar{e}, \alpha), (W, V, \Delta, u)$ is an optimal dual pair under r, θ, N* if and only if:

i) $q, s, \bar{e}, \alpha$ solves Problem 48

ii) for very x, ij, m, $\Delta_{ij}^m(x)$, $W_i^{out}(x)$, $W_i^m(x)$, $V_i(x)$ and u are the optimal Lagrange multipliers associated with constraints 37, 38, 39, 40, and 41, respectively.

iii) $Nu = f(r, q, s, \bar{e}, \alpha)$

Proof. Notice that Problem 48 is strictly convex and differentiable, hence its solution is characterized by the first order conditions. It is easy to see that these are given by Equations 43-47. In particular:

- Equation 45 is the first order condition with respect to $\mu(x)q_{ij}^m(x)$
- Equations 44 and 46 are equivalent to the first order condition with respect to $\mu(x)s_{i,h}^m(x)$
- Equation 43 is the first order condition with respect to $\mu(x)\bar{e}_i(x)$
- Equation 47 is the first order condition with respect to $\alpha_i^m(x)$

Hence $(q, s, \bar{e}, \alpha), (W, V, \Delta, u)$ is an optimal dual pair if and only if i) and ii) hold. To show that this implies iii) notice that summing the complementary slackness conditions of Constraints 37-40 and using

Constraint 41 gives

$$\begin{aligned} Nu = &\sum_x \mu(x)\{\sum_{m,ij} q_{ij}^m(x)[r_{ij}^m(x) - c^d d_{ij} - t_{ij}\nu(x)] - \sum_{m,i,h} s_{i,h}^m(x)\nu(x)\} \\ &- \sum_{x,i,m} \mu(x) \sum_h s_{i,h}^m(x) \frac{\mathrm{d}G^{\mathrm{d}}(s_i^m(x))}{\mathrm{d}s_{i,h}^m(x)} - \sum_{x,i} \frac{1}{\chi_x} \sum_m \alpha_i^m(x) \frac{\mathrm{d}G^{\mathrm{e}}(\alpha(x))}{\mathrm{d}\alpha_i^m(x)} \end{aligned}$$

From the results of Section A.1.1, this is equivalent to $Nu = f(r, q, s, \bar{e}, \alpha)$, completing the proof. □

A.2 Proof of Theorem 1

In what follows take r^{-m}, θ^{-m} as given. Note that, on the demand side, every $q^m, \bar{s}^m > 0$ can be implementing by setting customer prices according to Equation 18. Hence in this proof I focus on the supply side only. Notice that r^m, θ^m implements $q^m, \bar{s}^m$ if and only if there exists $q^{-m}, s, \bar{e}, \alpha, N$ and W, V, Δ such that $(q, s, \bar{e}, \alpha), (W, V, \Delta, \bar{u})$ is an optimal dual pair under r, θ, N. If this is the case for some $r^m, \theta^m, W, V, \Delta$, I will say that $q^{-m}, s, \bar{e}, \alpha, N$ *supports* $q^m, \bar{s}^m$. Define

$$\begin{aligned} c(r^{-m}, q, s, \bar{e}, \alpha, N) = &\sum_x \mu(x)\{\sum_{m',ij} q_{ij}^{m'}(x)[c^d d_{ij} + t_{ij}\nu(x)] + \sum_{m',i,h} s_{i,h}^{m'}(x)\nu(x)\} \\ &+ \sum_{x,i,m'} \mu(x) G^{\mathrm{d}}(s_i^{m'}(x)) + \sum_{x,i} \frac{1}{\chi_x} G^{\mathrm{e}}(\alpha_i(x)) + N\bar{u} \\ &- \sum_x \mu(x) \sum_{m' \neq m,ij} q_{ij}^{m'}(x) r_{ij}^{m'}(x) \end{aligned}$$

From Theorem 3 it follows that, at optimum, we must have

$$\sum_x \mu(x) \sum_{ij} q_{ij}^m(x) r_{ij}^m(x) = c(r^{-m}, q, s, \bar{e}, \alpha, N)$$

Intuitively, platform m must compensate drivers for their residual costs after taking into account the payments already received from other platforms. Hence the m's expenses associated with $q^m, \bar{s}^m$ are

pinned down by the tuple $q^{-m}, s, \bar{e}, \alpha, N$ supporting it, if any. Define the cost function

$$C(q^m, \bar{s}^m | r^{-m}, \theta^{-m}) = \min_{q^{-m}, s, \bar{e}, \alpha, N} c(r^{-m}, q, s, \bar{e}, \alpha, N) \tag{49}$$

$$\text{s.t.} 38 - 41$$

$$\mu(x) q_{ij}^{m'}(x) = \mu(x) \sum_h s_{i,h}^{m'}(x) \theta_{ij}^{m'}(x) \text{ for every } x, ij \text{ and } m' \neq m \tag{50}$$

$$\bar{s}_i^m(x) = \sum_h s_{i,h}^m(x) \text{ for every } x, i \tag{51}$$

Notice that, by definition of $\mathcal{L}(\theta^{-m})$, Problem 49 is feasible whenever $q^m, \bar{s}^m \in \mathcal{L}(\theta^{-m})$. Moreover, $q^{-m}, s, \bar{e}, \alpha, N$ supports $q^m, \bar{s}^m$ if and only if it solves Problem 49. To see this, for every x, ij, let $\Delta_i^m(x)$ be the optimal Lagrange multiplier associated with Constraint 51, and define $\theta_{ij}^m(x) \equiv q_{ij}^m(x)/\bar{s}_i^m(x)$, $\Delta_{ij}^m(x) = \Delta_i^m(x)/\theta_i^m(x)$, and

$$\begin{aligned} r_{ij}^m(x) = {} & \Delta_{ij}^m(x) + c^d d_{ij} + t_{ij}(\nu(x) + u) \\ & - t_{ij} \sum_{x'} \chi_{x,x'} V_j(x') - (1 - t_{ij}\chi_x) W_j^m(x) - W_i^m(x) \end{aligned}$$

For very x, ij, m', let $W_i^{out}(x)$, $W_i^m(x)$, $V_i(x)$ be the optimal Lagrange multipliers associated with constraints 37, 38, 39, and 40, and if $m' \neq m$, let $\Delta_{ij}^{m'}(x)$ be the optimal Lagrange multiplier of Constraint 50. Proceeding as in the proof of Theorem 3, it is easy to see that the optimality conditions of Problem Problem 48 hold. In particular, the optimality conditions of Problem 48 with respect to $q^{-m}, s^{-m}, \bar{e}, \alpha$ are identical to those of Problem 49, the optimality conditions with respect to q^m hold by definition of r, while the optimality conditions of Problem 49 with respect to s^m and N are given by

$$\begin{aligned} \frac{\mathrm{d}G^{\mathrm{d}}(s_i^m(x))}{\mathrm{d}s_{i,h}^m(x)} = {} & -\nu(x) - u + \Delta_i^m(x) \\ & + \sum_{x'} \chi_{x,x'} [V_i(x') - W_i^m(x)] + \sum_j \tau_{h,ij} [W_j^m(x) - W_i^m(x)] \end{aligned}$$

and

$$u = \bar{u}$$

respectively. This implies that every $q^m, \bar{s}^m \in \mathcal{L}(\theta^{-m})$ is supported by some $q^{-m}, s, \bar{e}, \alpha, N$, and for any such $q^{-m}, s, \bar{e}, \alpha, N$ we have $c(r^{-m}, q, s, \bar{e}, \alpha, N) = C(q^m, \bar{s}^m | r^{-m}, \theta^{-m})$. This proves the first statement of Theorem 1, while the second statement follows by the Envelope Theorem.

A.3 Proof of Theorem 2

Notice that, substituting 37 into 38-41, constraints 37-41 are equivalent to the following

$$\begin{aligned}
&\mu(x)\bar{e}_i(x)\chi_x = \alpha_i^{out}(x) \qquad (52)\\
&\mu(x)\sum_h s_{i,h}^m(x)[\theta_i^m(x) + \sum_j \tau_{h,ij} + \chi_x]\\
&\quad = \mu(x)\sum_{j,h} s_{j,h}^m(x)[\tau_{h,ji} + \theta_{ji}^m(x)(1-\chi_x t_{ji})] + \alpha_i^m(x)\\
&\sum_m \alpha_i^m(x) = \sum_{x'} \mu(x')[\sum_{m,h} s_{i,h}^m(x') + \sum_{m,j,h} s_{j,h}^m(x')\theta_{ji}^m(x')t_{ji} + \bar{e}_i(x')]\chi_{x',x}\\
&\sum_x \mu(x)[\sum_{m,i,h} s_{i,h}^m(x) + \sum_{m,ij,h} s_{i,h}^m(x)\theta_{ij}^m(x)t_{ij} + \sum_i \bar{e}_i(x)] = N
\end{aligned}$$

For every $\bar{s}$ define

$$\begin{aligned}
F(\bar{s}) = \min_{s,\bar{e},\alpha} &\sum_{x,i,m} \mu(x) G^{\mathrm{d}}(s_i^m(x)) + \sum_{x,i} \frac{1}{\chi_x} G^{\mathrm{e}}(\alpha_i(x))\\
&\text{s.t. } 52\\
&\mu(x)\bar{s}_i^m(x) = \mu(x)\sum_h s_{i,h}^m(x)
\end{aligned}$$

where I use the convention that $F^*(\bar{s}) = +\infty$ if the above problem is not feasible, and let F be its convex conjugate, defined by

$$F(\delta) = \max_{\bar{s}} [\sum_x \mu(x) \sum_{i,m} \bar{s}_i^m(x)\delta_i^m(x) - F^*(\bar{s})] \qquad (53)$$

Function $F(\delta)$ is strictly convex and differentiable.[42] By the Fenchel's duality Theorem, the following are equivalent:[43]

i) $\bar{s}$ solves Problem 53

[42] See for instance Galichon [2018, Proposition D.14]
[43] See for instance Galichon [2018, Proposition D.13]

ii) δ solves

$$\min_{\delta}[F(\delta) - \sum_{x} \mu(x) \sum_{i,m} \bar{s}_i^m(x)\delta_i^m(x)]$$

iii) $\bar{s} = \nabla F(\delta)$

Notice that, under the definition of δ in Equation 26, Problem 53 is equivalent to Problem 48. The result then follows from Theorem 3.

B Algorithms

B.1 Drivers' Problem

Rosaia [2020] provides an algorithm to compute undiscounted dynamic discrete choice models in discrete time. In this section I show how drivers' problem is equivalent to an analogous problem in discrete time, so that a discrete-time algorithm can be applied. Recall the drivers' value functions under δ, θ are characterized by the following system of equations

$$\begin{aligned}
W_i^{out}(x) &= -\frac{u}{\chi_x} + \sum_{x'} \frac{\chi_{x,x'}}{\chi_x} V_i(x') \\
U_{i,h}^m(x) &= \frac{\delta_i^m(x) - u}{\lambda_i^m(x)} + \sum_j \frac{\theta_{ij}^m(x)}{\lambda_i^m(x)}[-t_{ij}\bar{u} + t_{ij}\sum_{x'} \chi_{x,x'} V_j(x') + (1 - t_{ij}\chi_x) W_j^m(x)] \\
&\quad + \sum_j \frac{\tau_{h,ij}}{\lambda_i^m(x)} W_j^m(x) + \frac{\lambda - \sum_j \tau_{h,ij}}{\lambda_i^m(x)} W_i^m(x) + \sum_{x'} \frac{\chi_{x,x'}}{\lambda_i^m(x)} V_i(x') \\
V_i(x) &= \mathbb{E} \max_{m \in M \cup \{out\}} [W_i^m(x) + \frac{\eta_m}{\chi_x}] \\
W_i^m(x) &= \mathbb{E} \max_{h \in H} [U_{i,h}^m(x) + \frac{\epsilon_h}{\theta_i^m(x) + \lambda + \chi_x}]
\end{aligned} \tag{54}$$

where

$$\lambda_i^m(x) = \lambda + \theta_i^m(x) + \chi_x$$

Let

$$\phi = \max_{x,i,m} \max\{\frac{\lambda_i^m(x)}{1 + \sum_j \theta_{ij}^m(x) t_{ij}}, \chi_x\}$$

and for every x, ij, m define $\tilde{u} \equiv \frac{u}{\phi}$ and

$$\delta_i^m(x), \tilde{\theta}_{ij}^m(x), \tilde{\tau}_{h,ij}, \tilde{\lambda}, \tilde{\chi}_{x,x'} \equiv \frac{\delta_i^m(x), \theta_{ij}^m(x), \tau_{h,ij}, \lambda, \chi_{x,x'}}{\phi[1 + \sum_j \theta_{ij}^m(x) t_{ij}]}$$

Notice that Equations 54 can be re-written as

$$\begin{aligned}
W_i^{out}(x) &= -\tilde{u} + \sum_{x'} \frac{\chi_{x,x'}}{\phi} V_i(x') + (1 - \frac{\chi_x}{\phi}) W_i^{out}(x) \\
U_{i,h}^m(x) &= -\tilde{u} + \tilde{\delta}_i^m(x) + \sum_j \tilde{\theta}_{ij}^m(x)[t_{ij} \sum_{x'} \chi_{x,x'} V_j(x') + (1 - t_{ij}\chi_x) W_j^m(x)] \\
&\quad + \sum_j \tilde{\tau}_{h,ij} W_j^m(x) + (\tilde{\lambda} - \sum_j \tilde{\tau}_{h,ij}) W_i^m(x) + \sum_{x'} \tilde{\chi}_{x,x'} V_i(x') + [1 - \tilde{\lambda}_i^m(x)] U_{i,h}^m(x) \\
V_i(x) &= \mathbb{E} \max_{m \in M \cup \{out\}} [W_i^m(x) + \frac{\eta_m}{\chi_x}] \\
W_i^m(x) &= \mathbb{E} \max_{h \in H} [U_{i,h}^m(x) + \frac{\epsilon_h}{\theta_i^m(x) + \lambda + \chi_x}]
\end{aligned}$$

The above are the Bellman equations of an undiscounted dynamic discrete choice problem in discrete time. Hence $W, V, U, \tilde{u}$ can be computed by exploiting the algorithm in Rosaia [2020]. Given these, u can be obtained from $u = \phi\tilde{u}$.

B.2 Equilibria

This section describes the algorithm I use to find Platform Equilibria. The idea is very simple, and it can be described as a simultaneous gradient descent across platforms.

Algorithm. *Start from a transport equilibrium $\mathcal{P}, \mathcal{E}, N, \mathcal{A}$*

1. Compute δ from $\mathcal{P}, \mathcal{E}$ as in Equation 26, an compute the drivers' value functions associated with δ, θ

2. For every $m \in M$, use these value functions and the characterization of the cost functions' gradient in Equations 19 and 20 to compute the gradient of platforms' objective functions in 21 with respect to $q^m, \bar{s}^m$

3. For every m, increase/decrease $q^m, \bar{s}^m$ in the direction given by the gradient by a small amount. Let the new allocation be $\mathcal{A}^{+1}$

4. *Compute the new match qualities* $\mathcal{E}^{+1}$ *consistent with* $\mathcal{A}^{+1}$

5. *Find* $\mathcal{P}^{+1}, N^{+1}$ *such that* $\mathcal{P}^{+1}, \mathcal{E}^{+1}, N^{+1}, \mathcal{A}^{+1}$ *is a Transport Equilibrium*

6. *Continue until the gradients of platforms' objectives are close enough to zero*

Hence the algorithm stops at a Transport Equilibrium where the first order conditions of platforms' best response problem are all satisfied, that is, a Platform Equilibrium. The only step which is still unexplained is 5. To perform this step first notice that, on the demand side, the prices implementing $q, \bar{s}$ can simply be obtained from the inverse demand curves in 18. On the supply side notice that, for given N, we can compute the mean payoffs $\delta = \boldsymbol{\delta}(\bar{s})$ such that $\bar{s} = \bar{\boldsymbol{s}}(\delta)$ using the tâtonnement procedure described in Section 5.2.2, and compute the associated optimal average payoff flow $\boldsymbol{u}(\delta)$. Then N is increased (resp. decreased) by a small amount if $\boldsymbol{u}(\delta) >$ (resp. $<$)$\bar{u}$. This procedure can be iterated until an N is found under which $\boldsymbol{u}(\delta)$ is close enough to $\bar{u}$.