

# Spatial Distribution of Access to Service: Theory and Evidence from Ridesharing

Soheil Ghili, Vineet Kumar, Fei Teng\*

January 13, 2025

## Abstract

We study access to ridesharing across geographical regions, using both theoretical and empirical analyses. We specifically model and examine the effects of economies of density in ridesharing. Our model predicts that (i) economies of density skews access to ridesharing away from less dense regions, (ii) the skew will be more pronounced for smaller platforms (i.e., “thinner markets”), and (iii) ridesharing platforms do not find this skew efficient and thus use price and wage levers to mitigate (but not eliminate) it. We show that these insights are robust to whether the source of economies of density is the supply-side or the demand-side. We then calibrate our model using ride-level Uber data from New York City. We use the model to simulate counterfactual scenarios, offering a quantitative evaluation of our theoretical results and informing platform strategy and policy.

**Keywords:** Ridesharing; Spatial Markets; Transportation; Economies of Density; Market Thickness

## 1 Introduction

Ridesharing markets form a critical part of transportation networks in major metropolitan areas (Fortune Business Insights, 2021). In these markets, ridesharing firms act as two-sided platforms, intermediating between consumer (passenger) demand, and driver supply. Ridesharing markets are *spatial*, where the characteristics of demand and supply depend on location. Access to ridesharing service is an important issue that has drawn attention from regulators, lawmakers as well as consumer and community advocates (Diao et al., 2021; Jin et al., 2019). We focus on access, operationalized as the proportion of potential demand that is realized within a region. We examine how

---

\*Yale University. We thank Phil Haile, Igal Hendel, Nicole Immorlica, Vahideh Manshadi, Larry Samuelson, Yannis Stamatopoulos, and K. Sudhir. We also thank conference and seminar participants at Lyft Marketplace Labs, The ACM transactions on Economics and Computation (EC), Marketing Science, and Stanford Institute for Theoretical Economics (SITE) for helpful comments. We are thankful to Phonkrit Tavanisarut and Leonardo Fancello for excellent research assistance. All errors are our own.

market factors drive inequality in access to rides across different geographic areas within a large metropolitan city.

We use a combination of theoretical and empirical analysis to characterize access across regions. We first develop a theoretical model of a multi-region ridesharing market with a monopolist platform. We then take the model to data using trip-level data from New York City (NYC) focusing on the leading platform, Uber. Finally, we use the calibrated model parameters to evaluate counterfactual scenarios, which provide insights into the findings as well as the underlying mechanisms.

In the theoretical model, each region has an arrival rate of potential demand for rides that could end in the same region or a different one. Actual demand is a fraction of potential demand, depending on price (and pickup time) in the region. We incorporate drivers' and passengers' decision making among a set of regions. We model driver free entry, with each driver choosing whether and in which region to enter the market. Each driver makes this choice to maximize her revenue given other drivers' choices. Driver revenue in each region is positively related to the wage per ride in that region and negatively related to the "total wait time" each driver has to wait in the region to give a ride to a passenger. Total wait time consists of (i) "idle time," the time it takes for the driver to be assigned to a passenger requesting a ride, and (ii) "pickup time," the time it takes to arrive at the pickup location after being assigned to a passenger. Demand results from consumers choosing between a rideshare and an outside option, with price and pickup time factoring into their decision. The platform sets prices for consumers and wages for drivers in each region.

The elements of wait time play an important role in the resulting accesses in each region, and the differences across them. Both idle and pickup times are directly dependent on the number of drivers operating in a region. For a given demand, having more drivers operating in a region implies a higher expected idle time, since drivers are competing for the same pool of customers. The impact on pickup time is the inverse of the above. Having more drivers in a region implies a lower expected pickup time, since on average drivers are closer to passengers. Observe that while the idle time causes drivers to spread out across regions, the pickup time incentivizes drivers to agglomerate into fewer regions. We examine how the spatial distribution of access to rides is shaped in response to the incentives of the platform, drivers, and passengers. We consider two different sources of economies of density (EOD) or agglomeration: drivers directly preferring shorter pickup times (supply-side EOD) or demand responding positively to shorter pickup times (demand-side EOD).

The theoretical model's equilibrium analysis leads to some notable results. First, we find that access is skewed in equilibrium in favor of regions with higher density of demand. The underlying mechanism is that sparser regions have higher pickup times, so fewer drivers enter in equilibrium to balance this with idle times. Second, we find that platform size plays a significant role, with smaller platforms having more unequal access. Larger platforms have more demand, and correspondingly more driver entry, leading to lower pickup times in all regions. This reduces the importance of pickup times relative to idle times, resulting in more equal access across regions. Third, we find that a profit-maximizing ridesharing platform would optimally use price and wage levers to mitigate

but not fully eliminate the access skew. The platform implements this by offering higher wages to drivers in sparser regions but only partially passing the extra wage on to passengers, effectively subsidizing sparser regions. The reason is that driver entry into sparser regions has a greater positive externality on other drivers than entry into denser regions, which is internalized by the platform but not the drivers. Finally, we note that these results hold for both sources of economies of density, i.e. supply-side and demand-side.

We next take the theoretical model to data using data from Uber, the largest ridesharing platform in NYC, from March to June 2019. We leverage ride-level data on rides (including pickup and drop-off locations), driver wages, and prices across the regions (boroughs) of NYC. The dataset is publicly available from NYC’s Taxi and Limousine Commission (TLC), and is rich in frequency of observations, with the complete rides within and across the regions of the city available and collected in a standardized way. We use this data to calibrate and quantify the parameters of our model, e.g. unobservable potential demand, price sensitivity. We consistently find that the density of the potential demand is highest in Manhattan and lowest in Queens. Our estimates for measures such as the number of drivers and price elasticity are consistent with regulatory data and existing literature. We test a spectrum of models with different combinations of the supply-side and the demand-side EOD, and find these results to be robust.

We use our calibrated model to simulate several counterfactual scenarios of interest to complement our theoretical analysis. First, we simulate the platform’s optimal heterogeneous price and wage policy and find that in equilibrium, access in the most dense region (Manhattan) is higher by 25% than the least dense region (Queens). Next, we restrict the platform to using the same price and wage across regions. We find that under such a counterfactual, geographical skew in access to rides gets exacerbated – access in the most dense region is higher by 55% than the least dense region. These results indicate that the platform would use spatially differentiated pricing to mitigate the access inequality. The implication is that imposing equality in actions taken by the platform across regions could well exacerbate inequality in access that policy makers might care about. Finally, we assess the impact of platform size (i.e., market thickness) on outcomes by scaling up potential demand across all regions. Consistent with our theoretical predictions, we find that as platform size increases, driver entry rises, access levels improve in all regions, and access inequality diminishes. For instance, in one specification, when Uber’s potential demand grows from 80% to 120% of the estimated size, the access disparity between the most dense and least dense regions decreases, with the densest region being 28% denser at 80% demand and 22% denser at 120% demand, a 21% decrease in access disparity.

These results are of relevance to platforms, demonstrating the value of different strategies used, specifically decoupled price and wage levers across regions. The findings are also relevant to policy makers, who are considering a few ways to regulate platforms, which will impact access disparity. First, regulations imposing a standard wage rate or price across different regions (Parrott and Reich, 2018) can hurt access equity by not allowing the platform flexibility to incentivize drivers to serve sparser regions. Second, we consider a policy maker who wants to go the extreme case of

trying to equalize access across regions. Our model suggests that contrary to the optimal pricing policy of the platform (where prices are higher in sparser regions), prices for equalized access would need to be lower in sparser regions. Wages would also need to be substantially higher in sparser regions, leading to more pronounced wage difference across regions. Third, a regulation capping the number of drivers on a platform that has been tried is to reduce platform size (CBS News, 2018). Our model shows that imposing such a cap, while seeming to promote equality of outcomes across firms or higher pay for drivers, can actually make access disparity across regions worse.

In summary, our paper contributes to the study of access differences and economies of density across regions in ridesharing markets along several aspects. First, our model endows each region with a non-trivial size which allows to capture the notion of pickup time in each area and how it varies with driver density. This, in turn, allows us to model economies of density both on the supply side and on the demand side, unlike prior research. Second, our empirical strategy to recover access differences across regions based on relative flows of cross-region rides is powerful and generalizable to other passenger-transportation markets for inferring unfulfilled demand. This approach helps infer unobservable access levels using only ride-level data, which is what is commonly available in most markets, without requiring data based on individual drivers or consumers. Third, our model accommodates and characterizes economies of density based on region characteristics, driver behavior as well as strategic choices made by the platform that contribute to access differences. Finally, we extensively show (both theoretically and empirically) that the major implications of economies of density for ridesharing markets—as summarized in our main results—are robust to what proportion of economies of density arises from the supply side and what portion from the demand side.

There are a few aspects of our model that generalize beyond ridesharing to spatial markets more broadly. First, our results point to the fact that imposing constraints on firm actions (e.g., prices and wages) to make them more equal across regions may backfire by making other outcomes (e.g., access) more unequal across regions. Second, the concept of relative outflows across regions is generalizable to other passenger-transportation markets for inferring unfulfilled demand. Third, the idea that economies of density create inequity in spatial distribution is likely to exist across many other markets.

## 2 Literature Review

Our paper relates to multiple strands of literature: (i) the recent and growing literature on the empirical analysis of geographical distribution of supply, and its possible distortion from that of demand in spatial markets; (ii) the literature on transportation markets (in particular ridesharing); and (iii) the literature on the effects of market thickness in two-sided markets.

The empirical literature on spatial supply-demand matching is relatively new and limited. To our knowledge, only Buchholz (2022) and Brancaccio et al. (2023) directly examine this issue. These studies extend the empirical methods in the matching literature techniques (see Petrongolo

and Pissarides (2001) for a survey) to infer the size of unobserved demand (e.g., passengers searching for rides) in different locations of a decentralized-matching market when only the size of supply (e.g., available drivers) and the number of demand-supply matches (e.g., realized rides) are observed. Our calibration based on relative-outflows complements this approach. On the one hand, our method requires the additional assumption that potential demand for rides from region  $i$  to region  $i'$  is the same as that for rides in the opposite direction, which is plausible for passengers with a home base over a daily period. On the other hand, unlike prior methods, ours (i) requires only ride data rather than vacant supply; (ii) applies broadly to centralized (e.g., ridesharing) and decentralized (e.g., taxis) passenger-transportation markets; and (iii) detects supply skew from a given region even when passengers stop searching due to low supply, which might otherwise appear as low demand.

The second strand of literature relevant to our paper focuses on the market performance and design of transportation, particularly ridesharing, markets. While some studies focus on non-spatial questions (e.g., Cohen et al. (2016); Nikzad (2018)), others explore spatial aspects (such as Castillo et al. (2024); Frechette et al. (2019)) without examining the spatial distribution of supply and potential mismatches with demand. Research on geographical supply-demand (im)balances (e.g., Banerjee et al. (2018); Afèche et al. (2023); Besbes et al. (2021)) focuses on short-run, intra-day aspects. In contrast, some other papers (such as Bimpikis et al. (2019); Buchholz (2022); Shapiro (2018)) take a long-term view. Our paper complements this literature by providing a detailed theoretical and empirical analysis of economies of density (arising from both supply and demand sides) and market thickness, abstracting away from some of the phenomena considered in these papers. It is worth noting that a significant portion of this literature explores how ridesharing platforms improve upon traditional taxi systems, particularly through flexible pricing and advanced matching algorithms (Cramer and Krueger (2016); Cohen et al. (2016); Shapiro (2018); Castillo et al. (2024); Besbes et al. (2021); Lam and Liu (2017)). We contribute to this body of work by showing that, even within centrally matched ridesharing systems, platform size impacts driver incentives and access disparity across regions.

Also relevant to our work is Rosaia (2023), which examines the NYC ridesharing market using comprehensive data from Uber and Lyft and employs exogenous price shocks to model drivers' dynamic decision-making under platform competition. Compared to Rosaia (2023), our focus is on studying access inequity, and our approach uses a model that expands on the geographic aspects, focusing on the economies of density and market thickness, but abstracting from competition. We derive theoretical and empirical insights into these spatial implications, analyzing a broader geographical market that characterizes dense and sparse regions, empirically measuring access differences using relative flows, and conducting counterfactual analyses to inform firm strategy and policy.<sup>1</sup>

The third set of papers to which we relate is a large, mostly theoretical, literature on the impact of market thickness on the functioning of two-sided platforms in general (such as Akbarpour et al. (2020); Ashlagi et al. (2019)) and transportation markets in particular (such as Frechette et al.

---

<sup>1</sup>For a comparison of the geographical scope of the data used in our paper and Rosaia (2023), see Appendix J.

(2019); Nikzad (2018)). This literature, to our knowledge, has not examined how the spatial distribution of supply—and its (mis)alignment with that of potential demand—responds to a change in market thickness. Our paper focuses on this, both empirically and theoretically.

### 3 Theoretical Model

We develop a model to study the spatial distribution of access to ridesharing services in the presence of economies of density. We aim to use the model to understand (i) whether access will be skewed in favor of some regions over others in equilibrium, (ii) whether the skew will be exacerbated or moderated as the platform grows in size, (iii) whether the platform would like to use prices and wages as levers to influence the skew of access, and if so, the direction of this influence, and (iv) whether the answers to the previous questions are sensitive to whether economies of density arises from supply or demand side of the market. We abstract away from aspects of the ridesharing market that are not directly relevant to answering the above questions.

We develop a three-stage model where a monopolistic ridesharing platform optimally sets prices and wages, drivers decide whether to operate, and which region to operate in, and passengers decide whether to take the ridesharing service. We start with a setting without inter-region rides across regions (Section 3.1) and then introduce two variants of the model. First, we model the drivers' equilibrium spatial distribution where the economies of density comes only from the supply-side (referred to as the “model with supply-side EOD” or “supply-side model” for brevity) in Section 3.2. That is, drivers benefit from a larger number of drivers on the platform, as it reduces the pickup time and thus increases revenue. Second, in Section 3.3, we model the equilibrium spatial distribution of drivers with only demand-side economies of density, where we assume that ride demand decreases with pickup time and increases with the number of drivers (referred to as the “model with demand-side EOD” or the “demand-side model” for brevity).<sup>2</sup>

We prove the following set of results hold for both models: (i) under the platform optimal strategy, the equilibrium spatial distribution of rides is skewed towards regions with higher demand densities, (ii) the skew is more intensified for smaller ridesharing platforms, and (iii) the platform's optimal strategy would involve charging lower margins for rides in regions with lower demand densities. Then we extend the results to the setting with inter-region rides across regions in the supply-side model and the demand-side model respectively in Section 3.4.

#### 3.1 Model Setup

We model a market with regions  $i \in \{1, \dots, I\}$  and a monopolist ridesharing platform serving them. The regions (e.g. neighborhoods, boroughs) are modeled as circumferences of circles, a la Salop. The time that it takes a driver to travel a full circumference of circle  $i$  is denoted  $t'_i$ . Let  $t'$  denote

---

<sup>2</sup>Nevertheless, in both the supply-side model and the demand-side model, the agglomeration of drivers in a region increases the idle time and “turns away” drivers if the region becomes too dense. Therefore, under the assumption of a pool of infinite drivers and free entry of drivers, the equilibrium spatial distribution of drivers is reached when the two forces, the pickup time and the idle time, are balanced.

the vector  $(t'_1, \dots, t'_I)$ . For ease of exposition, we assume  $t'_i$  is measured in hours. Intuitively,  $t'_i$  captures the geographical size of the region. For any given numbers of drivers and riders, the drivers and the riders distribute more sparsely in a region as  $t'_i$  becomes larger.

In each region, the platform decides price per ride  $p_i$  and wage per ride  $c_i$ .<sup>3</sup> Let  $p$  denote the vector of  $(p_1, \dots, p_I)$  and  $c$  denote the vector of  $(c_1, \dots, c_I)$ . A distribution of drivers is denoted by vector  $n = (n_1, \dots, n_I)$ , where  $n_i$  denotes the number of drivers in region  $i$ . To ease the analysis, we assume  $n_i$  are real numbers as opposed to integers. Let  $\omega_i$  denote the pickup time of each ride in region  $i$ . Passengers arrive at a rate  $\lambda_i$  per unit of time (i.e.  $\lambda_i$  passengers arrive per hour). Demand arrival rate is a function of price per ride  $p_i$  and can also be a function of  $\omega_i$ , depending on our assumption of the source of economies of density. We discuss this in further details in the subsequent sections. For now, we assume that the demand arrival rate takes the following form:

$$\lambda_i(p_i, \omega_i) = \bar{\lambda}_i f(p_i, \omega_i)$$

$\bar{\lambda}_i$  is the “potential demand” in region  $i$ , which captures the total demand of transportation in region  $i$  per hour and thus is the maximum number of realized rides achievable in region  $i$  for the ridesharing platform. Equivalently, this is the volume of demand when both  $p_i$  and  $\omega_i$  are zero. Function  $f(\cdot)$ —which is assumed uniform across regions—captures the fraction of  $\bar{\lambda}_i$  that would be willing to pay  $p_i$  for a ride given they observe pickup time  $\omega_i$ , which is the time it takes a driver, after being assigned to a ride request, to drive and arrive at the passenger’s pickup location.  $f(\cdot)$  takes the values between 0 and 1.  $\bar{\lambda}$  represents the vector  $(\bar{\lambda}_1, \dots, \bar{\lambda}_I)$ .

Given  $t'_i$ ,  $\lambda_i$  and  $n_i$ , there is a wait time for drivers before they can provide a ride to a passenger. This total wait time is denoted by  $W_i(n_i)$  and is a function of  $n_i$ , the number of drivers present in the region (this notation suppresses the implicit dependence of  $W_i$  on prices, as we detail later). Total wait time in each region has two components: idle time and pickup time. Idle time is the time it takes a driver to get assigned to a ride requested by a passenger. Idle time in region  $i$  is increasing in  $n_i$ . That is, the more drivers in region  $i$ , the longer it takes for each of them to get assigned to a passenger. Pickup time  $\omega_i$  in region  $i$  is decreasing in  $n_i$ . This is because the more drivers in region  $i$ , the more densely the region is populated with them. Therefore, each driver becomes less likely to be asked to pick up a passenger who is far away in the region.<sup>4</sup> We assume that each arriving passenger’s location is uniformly distributed on the circumference of the circle. Then the pickup time  $\omega_i$  as a function of  $n_i$  can be written as

$$\omega_i(n_i; t'_i) = \frac{t'_i}{4n_i}$$

The total wait time for each region  $i$  is given by:

---

<sup>3</sup>We assume that the rides are homogeneous.

<sup>4</sup>We assume that the centralized platform assigns each passenger to the closest driver. See Appendix C for details.

$$\underbrace{W_i(n_i)}_{\text{Total Wait Time}} = \underbrace{\frac{n_i}{\lambda_i}}_{\text{Idle Time}} + \underbrace{\frac{t'_i}{4n_i}}_{\text{Pickup Time}} = \left( \frac{n_i}{\lambda_i} + \frac{t'_i}{n_i} \right) \quad (1)$$

where  $t_i = \frac{t'_i}{4}$ . In Appendix C, we provide a micro-foundation for this functional form. From this point on, we refer to  $t_i$  (instead of  $t'_i$ ) as the size of region  $i$ . Let  $t$  denote the vector  $(t_1, \dots, t_I)$ . Given  $t_i$  and  $\bar{\lambda}_i$ , we define the density of potential demand for region  $i$  as  $\frac{\bar{\lambda}_i}{t_i}$ . Without loss of generality, we assume that the density of potential demand decreases in region index:  $\forall i < j : \frac{\bar{\lambda}_i}{t_i} \geq \frac{\bar{\lambda}_j}{t_j}$ .

We model a game of three stages. In the first stage, the platform optimally decides prices per ride  $p_i$  and wages per ride  $c_i$  for each region  $i$  to maximize the total profit across all regions. In the second stage, drivers from an infinitely large pool simultaneously decide whether to enter the market and, if so, which region to operate in. In the third stage, consumers observe both price  $p_i$  and pickup time  $\omega_i$  and decide whether to choose the ridesharing service. Then actual demand  $\lambda_i$  is realized.

We solve the equilibrium using backward induction starting from the third stage. In the third stage, given  $\bar{\lambda}_i$ ,  $p_i$  and  $\omega_i$ , the actual demand per hour is  $\lambda_i = \bar{\lambda}_i f(p_i, \omega_i)$ . In the second stage, each driver seeks to maximize her expected hourly revenue. The hourly revenue for a driver in each region  $i$  equals the wage per ride in that region multiplied by the frequency of rides assigned to each driver in the region. That is, the revenue is  $\frac{c_i}{W_i(n_i)}$ .<sup>5</sup> Let  $\bar{c}$  denote the exogenously given reservation value, i.e., the hourly revenue that each driver can make by leaving the market and taking the outside option.

Now we formally define the partial equilibrium distribution of drivers (given  $p_i$  and  $c_i$ ) as follows:

*Definition 1.* Under “market primitives”  $(\bar{\lambda}, t, \bar{c})$  and given the platform’s strategy  $(p, c)$ , a distribution of drivers  $n^* = (n_1^*, \dots, n_I^*)$  among the  $I$  regions is called an equilibrium if no small mass of drivers currently operating in region  $i$  can strictly profit from changing their strategy and operating instead in region  $j$ , where  $i, j \in I$ .<sup>6</sup> That is,  $\exists \delta > 0$  s.t.  $\forall i, j \in I, \forall \delta' \in [0, \delta]$  we have:<sup>7</sup>

$$\frac{c_i}{W_i(n_i^*; p_i)} \geq \frac{c_j}{W_j(n_j^* + \delta'; p_j)}$$

We call  $n^*$  an “all-regions” equilibrium allocation if it is an equilibrium and if  $n_i^* > 0$  for all  $i$ .

Directly following the definition above, we have the following statement that gives the necessary condition for the partial equilibrium of driver distribution:

---

<sup>5</sup>Note that this formulation abstracts from the time it takes to drive passengers to dropoff locations, simplifying some of our analysis. We do not expect the results to be qualitatively sensitive to it.

<sup>6</sup>The reason why we focus on “small mass” deviations is that our continuous-mass model of drivers is an approximation for a large population. In this environment, allowing for “large-mass” deviations from any given strategy should be interpreted as drivers *coordinating* a deviation, which we do not allow.

<sup>7</sup>Note that the wait time  $W_i$  is a valid definition only if  $n_i^* > 0$ . If the drivers are currently taking the outside option in region  $i$  (i.e.  $n_i^* = 0$ ), then the left hand side of the inequality is replaced by  $\bar{c}$ . If the drivers are considering switching to outside option, then the right hand side of the inequality is replaced by  $\bar{c}$ .

**Lemma 1.** Suppose function  $W_i(\cdot; \cdot)$  is continuous. For any given  $(p, c)$ , if  $n = (n_1^*, \dots, n_I^*)$  with  $n_i^* > 0 \forall i$  is the equilibrium distribution of drivers, then

$$\frac{c_i}{W_i(n_i^*; p_i)} = \bar{c}, \quad \forall i \quad (2)$$

Given the drivers' response, in the first stage of the game, the platform optimally chooses  $(p, c)$  based on market primitives  $(\bar{\lambda}, t, \bar{c})$  to maximize the platform profit, which is the sum of the regional profits. The total number of rides per hour in region  $i$ , denoted  $r_i(n_i)$ , is given by the total number of drivers in that region divided by the time each driver has to wait before giving a ride:  $r_i(n_i) \equiv \frac{n_i}{W_i(n_i)}$ . If the distribution of drivers satisfies eq. (2), then  $r_i(n_i^*) = \frac{n_i^*(p_i, c_i)}{c_i/\bar{c}}$  in equilibrium. The platform's profit per hour, which is the object the platform seeks to maximize by choosing regional prices and wages, is given by:

$$\pi(p, c) = \sum_{i=1}^I (p_i - c_i) r_i(n_i(p_i, c_i))$$

Then the equilibrium of the game is given by  $(p^*, c^*, n^*, \lambda^*)$ , where  $p^*$  and  $c^*$  are the prices and wages that maximize platform profit,  $n^*$  is the equilibrium driver distribution given  $(p^*, c^*)$ , and  $\lambda^*$  is the actual demand corresponding to prices  $p^*$  and pickup times  $\omega_i(n^*)$  of each region.

Next, we introduce the notion of “access” to rides in region  $i$  by  $A_i(n_i)$  and define it as the fraction of the potential demand  $\bar{\lambda}_i$  that leads to rides. That is:

$$A_i(n_i) \equiv \frac{r_i(n_i)}{\bar{\lambda}_i} \quad (3)$$

### 3.2 Supply-Side Economies of Density

In this section, we assume that the market involves supply-side economies of density (but does not involve demand-side economies of density). That is, a driver in a given region  $i$  can benefit from the presence of other drivers in that region because with higher driver density, she is less likely to be asked to pick up a passenger far away in the region. On the other hand, to abstract from the effect from the demand-side EOD, we assume that passengers' arrival rate does not depend on the number of drivers (and hence on pickup time). Specifically, given  $\bar{\lambda}_i$ , the demand arrival rate of region  $i$  is only a function of  $p_i$ , and we assume that  $f(p_i, \omega_i) = 1 - \alpha p_i$  so that <sup>8</sup>

$$\lambda_i = \bar{\lambda}_i(1 - \alpha p_i)$$

---

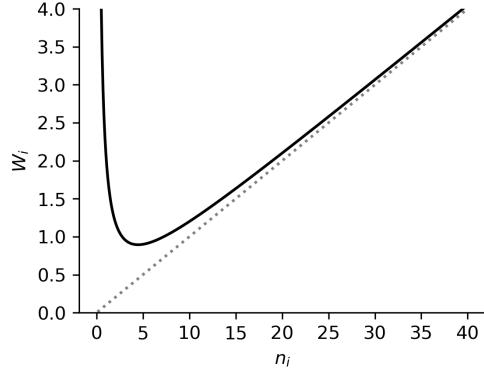
<sup>8</sup>The linear demand model has been used in literature. See e.g., Bimpikis et al. (2019). We also provide a micro-foundation here. The demand function can be derived under the following assumptions: (1) there is one unit of potential customers and (2) the valuation for each ride in region  $i$ , denoted  $v_i$ , follows a uniform distribution  $unif(0, \frac{1}{\alpha})$ . That is, at a given price  $p_i$ , only customers with  $v_i > p_i$  will demand the ride from the platform, resulting in  $1 - \alpha p_i$  of the population choosing to ride.

Accordingly, the wait time is given by:

$$W_i(n_i) = \frac{n_i}{\bar{\lambda}_i(1 - \alpha p_i)} + \frac{t_i}{n_i}, \quad (4)$$

where the first component is the idle time and the second component is the pickup time. The wait time curve given by eq. (4) is illustrated in Fig. 1. As can be seen there,  $W_i(\cdot)$  is initially decreasing in  $n_i$  because the effect of pickup time is dominant. When  $n_i$  is large enough, pickup time becomes less important and  $W_i(\cdot)$  becomes increasing in  $n_i$  due to the effect of idle time.

Figure 1: Wait Time as a Function of the Number of Drivers



Note. The function is given by given by eq. (4). The figure is illustrated with  $\lambda_i(p_i) = 10$  and  $t_i = 2$ . The dashed line is  $\frac{n_i}{\bar{\lambda}_i}$ , depicting how long the wait time would be if pickup time were zero.

In the propositions below, we characterize the equilibrium spatial distribution of supply, and how this distribution changes in response to a changed market thickness.

**Lemma 2.** *Suppose prices and wages are both positive and flexible and that market primitives are  $(\bar{\lambda}, t, \alpha, \bar{c})$ . Then an equilibrium  $(p^*, c^*, n^*)$  exists and  $n^*$  is unique. Also  $p_i^*$  and  $c_i^*$  for any  $i$  with  $n_i^* > 0$  are unique.*

**Proposition 1.** *Suppose prices and wages are both flexible and that market primitives are  $(\bar{\lambda}, t, \alpha, \bar{c})$ . Also suppose  $(p^*, c^*, n^*)$  is an equilibrium. Then for any two regions  $i, j$  with  $n_i^* \neq 0 \neq n_j^*$ , we have:*

$$\frac{\bar{\lambda}_i}{t_i} \geq \frac{\bar{\lambda}_j}{t_j} \Rightarrow A_i(n_i^*) \geq A_j(n_j^*)$$

where the latter comparison holds with equality only if the first one does. Moreover, regions that are not supplied are those with lowest demand densities:

$$\exists \mu > 0 : s.t. \forall i : \frac{\bar{\lambda}_i}{t_i} < \mu \Leftrightarrow n_i^* = 0$$

The proof is provided in Appendix E.1, but here we develop an informal intuition for the above proposition in two steps. First, we argue that if all prices and wages were spatially homogeneous,

equilibrium driver allocation (fixing those prices and wages) would give higher access to rides in denser regions relative to sparser ones. Next, we argue that the platform does not have the incentive to use prices and wages to fully eliminate such an access skew.

For intuition of the first argument in the previous paragraph, consider regions  $i$  and  $j$  with  $\frac{\bar{\lambda}_i}{t_i} > \frac{\bar{\lambda}_j}{t_j}$  and observe that with spatially homogeneous prices and wages, the total wait time for drivers should be the same between the two regions. This directly implies that the distribution of drivers should be skewed toward  $i$ , that is:  $\frac{n_i}{n_j} > \frac{\bar{\lambda}_i}{\bar{\lambda}_j}$ . To see why, note that if we instead had  $\frac{n_i}{n_j} = \frac{\bar{\lambda}_i}{\bar{\lambda}_j}$ , then idle times would be equal between the two regions but pickup time in region  $i$  would be shorter than that in  $j$ , incentivizing drivers to relocate from  $j$  to  $i$ . But given total driver wait times are equal between the regions,  $\frac{n_i}{n_j} > \frac{\bar{\lambda}_i}{\bar{\lambda}_j}$  is equivalent to  $\frac{r_i}{r_j} > \frac{\bar{\lambda}_i}{\bar{\lambda}_j}$  which by definition means  $A_i > A_j$ .

It is helpful to understand the intuition for why the platform does not benefit from fully eliminating the access gap. The simple argument here is that drivers spending long times en route for pickups hurts not only drivers themselves but also the profitability of the platform. As a result, it should not be surprising that the platform's optimal decision accounts for driver incentives (and pays drivers more to compensate for longer pickup times).

Consider the case of equal access across all regions. Drivers in less dense regions would have a greater pickup time, whereas the idle time would be the same across all regions if the price and the wage are the same. In such a case, drivers would have a higher wait time in sparser regions, and would have to be compensated more.

Next, we turn to examining the impact of platform size (i.e. market thickness) on the access skew across regions. Before that, we provide a formal definition of market thickening.

*Definition 2.* Consider a market with the primitive of potential demand as  $\bar{\lambda}$ . We call a market with potential demand  $\gamma\bar{\lambda}$  with  $\gamma > 1$  and all other primitives remaining the same a “thickening” of the market with potential demand  $\bar{\lambda}$ .

Intuitively, market thickening increases the potential demand in each region proportionally so that the ratio of potential demand is *preserved* between any two regions. Nevertheless, as our next result shows, making a market thicker will mitigate the skew of supply distribution.

**Proposition 2.** *Suppose prices and wages are both flexible and that market primitives are  $(\bar{\lambda}, t, \alpha, \bar{c})$ , for which  $(p^*, c^*, n^*)$  is an equilibrium. Consider a thickening of the market from  $(\bar{\lambda}, t, \alpha, \bar{c})$  to  $(\gamma\bar{\lambda}, t, \alpha, \bar{c})$  where  $\gamma > 1$ . Then, the following are true:*

1. *There exists an equilibrium  $(p^{*'}, c^{*'}, n^{*'})$  under the new primitives  $(\gamma\bar{\lambda}, t, \alpha, \bar{c})$ , where  $n^{*'}$  is unique. Also  $p_i^{*'} and  $c_i^{*'} for any  $i$  with  $n_i^{*'} > 0$  are unique.$$*
2. *For any  $i$ :  $n_i^* > 0 \Rightarrow n_i^{*'} > 0$ .*
3. *For any  $i, j$ :*

$$\frac{\bar{\lambda}_i}{t_i} \geq \frac{\bar{\lambda}_j}{t_j} \Rightarrow \frac{A_j(n_j^*)}{A_i(n_i^*)} \leq \frac{A_j(n_j^{*'})}{A_i(n_i^{*'})} \leq 1$$

4. There will be equitable access to rides as the market gets sufficiently thick:

$$\forall i, j : \lim_{\gamma \rightarrow \infty} \frac{A_j(n_j^{*\prime})}{A_i(n_i^{*\prime})} = 1$$

Recall we assume that passengers are matched with the closest drivers.<sup>9</sup> The intuition for the result is that as market thickens (i.e., the platform grows), all regions become denser with both drivers and passengers. As such, the importance of pickup times relative to idle times decreases in drivers' decision making, leading to a supply distribution that is more balanced with demand.

We now delve deeper into the incentives of the platform and how it aligns with those of the drivers. In interpreting Proposition 1, we argued that the platform, similarly to drivers, does suffer from long pickup times and is, hence, not willing to use prices and wages to fully eliminate access skew among regions of different densities. The next natural question is: does the platform have the incentive to at least mitigate the skew? The following proposition sheds light on this.

**Proposition 3.** *Suppose prices and wages are both flexible and that market primitives are  $(\bar{\lambda}, t, \alpha, \bar{c})$ . Also suppose  $(p^*, c^*, n^*)$  is an equilibrium. For any regions  $i, j$  with  $n_i^* \neq 0 \neq n_j^*$ , we have:*

$$\frac{\bar{\lambda}_i}{t_i} \geq \frac{\bar{\lambda}_j}{t_j} \Rightarrow \begin{cases} p_i^* \leq p_j^* \\ c_i^* \leq c_j^* \\ p_i^* - c_i^* \geq p_j^* - c_j^* \end{cases}$$

where the latter three comparisons hold with equality only if the first one does.

As this proposition shows, the platform obtains lower margins ( $p_i - c_i$ ) for its services in less dense regions. That is, the platform optimally gives higher wages to drivers in sparser regions but does not pass the entire wage increase on to passengers in the form of higher prices. Thus, the platform effectively “subsidizes” sparser regions, relative to denser ones. This is consistent with an incentive to mitigate the access skew across regions. The intuition for why the platform would like to mitigate the access skew has to do with *externalities*. Consider a driver's decision to operate in a denser region  $i$  as opposed to a sparser region  $j$ . This choice leaves region  $j$  even sparser, thereby increasing the pickup time in  $j$  and negatively affecting incentives of other drivers to choose  $j$ . While the driver's decision to operate in region  $i$  makes the region even denser, the positive marginal impact of region  $i$  being denser does not compensate for the loss from region  $j$  being sparser. This externality on other drivers is internalized by the platform through profits, but not directly internalized by the driver. This creates a wedge between platform and driver incentives and leads the platform to use price and wage levers to mitigate such externalities by drivers.

There are a few interesting observations. First, the platform's optimal strategy is more complex than what might seem based on the intuition in the above paragraph. The complexity arises from the fact that the platform simultaneously determines optimal prices and optimal wages: offering

---

<sup>9</sup>See Appendix C for details.

higher wages in sparser regions put some upward pressure on prices. Likewise, lower prices in sparser regions put some downward pressure on wages. One question, which Proposition 3 above helps answer, is “which lever, prices or wages, does the platform use to mitigate the access skew?” Proposition 3 suggests that in sparser regions, the platform optimally builds economies of density by providing higher wages. Despite the platform charging higher prices, which reduces the economies of density, overall the effect of lower wage dominates. Thus, the platform trades off getting a lower margin in sparser regions in order to build economies of density.

In the empirical section (Section 6.1), we connect the margin results in Proposition 3 to access. Specifically, in our counterfactual simulations, we show that access among regions becomes more skewed if the platform is restricted to spatially uniform prices and wages (and thus, spatially uniform margins). With our analysis of supply-side model completed, we now turn to a version of the model where economies of density arises only from the demand side; and we study if the above insights are robust to such a change in the model.

### 3.3 Demand-Side Economies of Density

The pickup time for a driver is equivalent to the total wait time experienced by a passenger riding with that driver, and decreases as the number of drivers increases. Therefore, it is natural to model the riders’ sensitivity to pickup time as demand-side EOD and examine the consequences of it. This would be a model for demand-side economies of density: higher pickup times in sparser regions diminish demand in those regions, making the market sparser and further increasing local pickup times.

We model demand-side EOD *only*, where pickup times do not directly impact driver behavior or revenue but impact passengers. To incorporate the latter assumption, we need to make demand  $\lambda_i$  not only a function of prices, but also pickup times:

$$\lambda_i(p_i, n_i) = \bar{\lambda}_i(1 - \alpha p_i) \left(1 - \beta \frac{t_i}{n_i}\right),$$

where  $\alpha > 0$  measures the price sensitivity and  $\beta > 0$  measures the sensitivity to pickup time, thereby capturing the magnitude of demand-side economies of density.

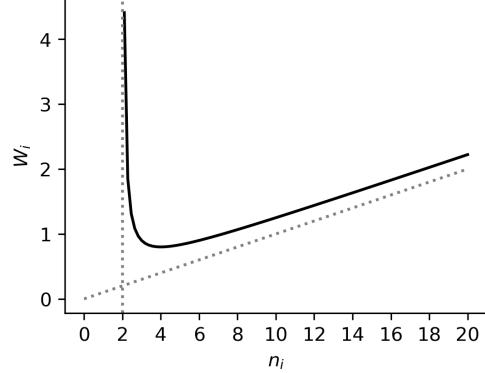
To capture the idea that pickup times do not directly impact drivers, we need to make two assumptions. First, region sizes  $t_i$  (and hence pickup times) are so small that driver wait time is approximated by idle time only. Second, in spite of all  $t_i$  values being vanishingly small, the sensitivity to pickup times  $\beta$  is sufficiently large so that  $\beta \times t_i$  values are non-negligible and, hence, passengers still care about pickup times. In this setting, driver wait times are given by:

$$W_i(n_i) = \frac{n_i}{\lambda_i(p_i, n_i)} = \frac{n_i}{\bar{\lambda}_i(1 - \alpha p_i)(1 - \beta \frac{t_i}{n_i})}, \quad (5)$$

The wait time curve is illustrated in Fig. 2. Similar to the model with supply-side EOD,  $W_i(\cdot)$  initially decreases in  $n_i$  as the effect of pickup time dominates. As  $n_i$  is large enough,  $W_i(\cdot)$  becomes

increasing in  $n_i$  due to the effect of idle time. The number of rides per hour,  $r_i$ , and access,  $A_i$ , are as previously defined.

Figure 2: Wait Time as a Function of the Number of Drivers



Note. The function is given by eq. (5). The figure is illustrated with  $\bar{\lambda}_i(1 - \alpha p_i) = 10$  and  $\beta t_i = 2$ .

The following proposition characterizes the equilibrium spatial distribution of supply and its response to market thickness under demand-side economies of density.

**Proposition 4.** *Suppose the market primitives are  $(\bar{\lambda}, t, \alpha, \beta, \bar{c})$ . Both prices and wages are flexible. Under the model with demand-side economies of density, all propositions with the model with supply-side economies of density (Lemma 2, Propositions 1 to 3) continue to hold.<sup>10</sup>*

The key lesson from this proposition is that our main insights are robust to the source of economies of density. This is because both the supply-side and demand-side models share a similar mechanism. We detail the underlying intuition connecting the models.<sup>11</sup> In equilibrium, drivers distribute across regions to balance the idle time and pick-up time in different regions. Recall that idle time increases with number of drivers in a region, whereas pick-up time asymptotically goes to zero as number of drivers grows. The difference between the supply-side and demand-side models lies in how the pick-up time affects drivers' payoffs. In the supply-side model, pick-up time directly affects drivers' payoffs since they internalize this as a cost, similar to idle time. In the demand-side model, drivers are not directly sensitive to pickup time. However, they are indirectly impacted by it, since consumer demand is responsive to pickup time (in addition to price). Specifically, demand decreases as pickup time increases. Thus, in both models, pickup time has either a direct or indirect impact on the drivers in the same direction. Since both direct and indirect effects above are directionally (and qualitatively) similar in terms of the impact on drivers, we should expect the main results to hold regardless of the source of economies of density (supply-side or demand-side). Proposition 4 verifies that this is indeed the case.

<sup>10</sup>One small exception is that in the demand-side model, the first comparison in Proposition 3 (i.e.,  $p_i^* \leq p_j^*$ ) always holds with equality.

<sup>11</sup>We thank an anonymous reviewer for this point.

In addition to being interesting in its own right, the robustness of our main insights to the source of economies of density will play a key role in our empirical analysis: there, we will argue that demand- and supply-side economies of density are not separately identifiable from each other (at least with ride-level data). Hence, robustness of our main insights to the source of economies of density becomes crucial when it comes to practical recommendations to firms and policymakers. Next, we include inter-region rides to make the model realistic for empirical analysis.

### 3.4 Inter-Region Rides

An abstraction in our model so far is the assumption that rides take place only within regions. In this section, we first define the equilibrium notion with inter-region rides. Then we show that Propositions 1 to 4 continue to hold with inter-region rides.

Let  $\bar{\lambda}_{ij}$  denote the potential demand for rides from region  $i$  to region  $j$ . By construction, we have  $\bar{\lambda}_i = \sum_j \bar{\lambda}_{ij}$ . Define realized demand  $\lambda_{ij}(p_i, \omega_i)$  in a similar way:  $\lambda_{ij}(p_i, \omega_i) = \bar{\lambda}_{ij} f(p_i, \omega_i)$ . Define  $r_{ij}$  as the realized number of rides from  $i$  to  $j$ . In this framework, one can think of  $n_i$ , the number of drivers in region  $i$ , as a stock variable with flows in and out. Realized  $r_{ij}$  values lead to “cross-regional flows” of drivers. These flows of drivers are already in the market. In addition, we also allow for a (negative or positive) net flow of drivers  $\rho_i$  from each region  $i$  to the outside option. In this setting, we need a new notion of equilibrium which is given by Definition 3.

*Definition 3.* Suppose the “market primitives”  $(\bar{\lambda}, t, \bar{c})$  and platform strategy  $(p, c)$  are given. Let  $\rho_i$  denote the (positive or negative) net flow of drivers out of each region  $i$ .  $n^* = (n_1^*, \dots, n_I^*)$  denote the distribution of drivers among the  $I$  regions. The pair of vectors  $(n^*, \rho^*)$  is an equilibrium if it satisfies the following two conditions:

1. **Steady State:** Net flow out of each region  $i$  is zero:  $\forall i : \rho_i^* + \sum_j r_{ij}^* - \sum_j r_{ji}^* = 0$
2. **Optimality:** No small perturbation in rates  $\rho^*$  is able to strictly improve driver revenue in any region. Formally, fixing some elapsed time  $\bar{T} > 0$ , and perturbation  $\Delta\rho \neq 0$ , the following will hold  $\forall i \in I, \forall T \in (0, \bar{T})$ :

$$\frac{c_i}{W_i(n_i^*)} \geq \frac{c_i}{W(n_i^* + T \times \Delta\rho)}$$

Also, we call  $n^*$  an “all-regions” equilibrium allocation if  $(n^*, \rho^*)$  it is an equilibrium and if  $n_i^* > 0$  for all  $i$ .

We now show that the results provided in the previous sections (i.e., without inter-region rides) all hold with inter-region rides as well. Proposition 5 formalizes this.

**Proposition 5.** *Suppose market primitives  $(\bar{\lambda}, t, \bar{c})$  and platform strategy  $(c, p)$  are given. Then for any  $(n^*, \rho^*)$  with  $n^* >> 0$ , the pair  $(n^*, \rho^*)$  is a driver equilibrium according to Definition 3 if and only if  $n^*$  is a driver equilibrium in the no-inter-region-ride version of the problem.*

The proof of this proposition is straightforward and is provided in Appendix E.3. This result shows that as long as we focus on all-region equilibria, all of our results from Section 3.2 and Section 3.3 should hold. The intuition is simple. If there are too many incoming rides into a region  $i$  with overall low payoff, that will lead to many drivers opting to exit the market and pursue an outside option upon dropping a passenger in  $i$ . Similarly, attractive payoff in a region  $i$  with low inflow of rides will garner higher entry of local drivers.

Note that the purpose of Proposition 5 is *not* to provide conceptually new insights on economies of density (the proposition only claims that the old results hold). The purpose is, rather, to connect our theory to the forthcoming empirical analysis, where we will use relative magnitudes of inter-region rides to identify the strength of economies of density and measure access skew. Next, we detail the data used in the empirical analysis.

## 4 Data and Summary Statistics

The primary source of data is the ride-level data on Uber in New York City (NYC) from March to June in 2019. The dataset is publicly available from NYC's Taxi and Limousine Commission (TLC). Uber is the largest ridesharing platform whose total number of rides given in NYC per month was about 14.4 million in June 2019. The regions are defined as the four boroughs of NYC: Bronx, Brooklyn, Manhattan and Queens.<sup>12</sup> We use the terms *region* and *borough* interchangeably in the following sections of this paper. For each ride, we observe the exact date, time, and location both for the pickup and the drop-off. Besides, we observe the wage paid to the driver and the price paid by the passenger.<sup>13</sup>

There are a number of reasons why we chose the NYC TLC Data from March to June 2019 as the setting for our empirical analysis. First, TLC provides data on both pickup and drop-off locations. Second, starting in February 2019, additional ride-level price and wage data became available. Both of these features play key roles in our analysis. Third, NYC is one of the densest cities in the world. Therefore, if frictions from low density impact the spatial distribution of supply and access in NYC, it is reasonable to conclude they must also be relevant in other ridesharing markets.

We aggregate the ride-level data at borough-day level. For each borough  $i$  and day  $d$ , we calculate the average price per ride,  $p_{id}$ , and the average wage per ride,  $c_{id}$ , for all rides originating from borough  $i$  on day  $d$ . Fig. 3 illustrates the daily average prices and wages by borough throughout the dataset. Additionally, we compute the average number of rides picked up from each borough per hour, denoted as  $r_{id}$ . Table 1 reports the summary statistics of these variables for each borough across days, revealing substantial variation in all three variables across regions and days. For the reservation level of wage,  $\bar{c}$ , set to \$24, we rely on figures from Hall and Krueger (2018), who report

---

<sup>12</sup>We exclude rides either picked up or dropped off in Staten Island because the rides between Staten Island and other boroughs are scarce on some of the days within our data period. This leads to extreme outliers in the measures of relative outflows and biases the estimates.

<sup>13</sup>We exclude shared rides, since the data does not provide separate wage and price for each co-rider. We retain rides with positive trip time and distance and exclude rides where the passenger fare is lower than the driver pay.

this as the mean hourly gross earnings before expenses for NYC UberX drivers for October 2015.<sup>14</sup> We set  $\bar{c}$  to be uniform across boroughs and across days. Finally, we use data on the borough areas (in square miles)<sup>15</sup> and calculate the square root of each area, denoted as  $s_i$ .<sup>16</sup>

Figure 3: Price and Wage across Days and Boroughs

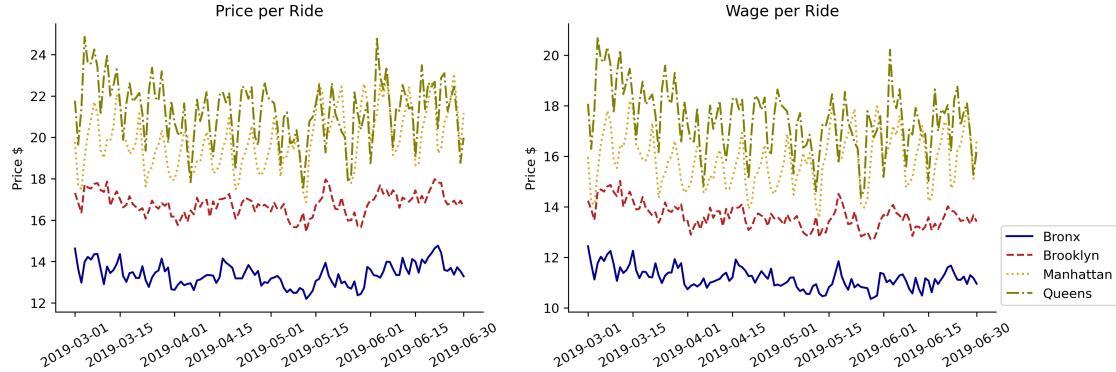


Table 1: Summary Statistics of Number of Rides, Price and Wage at Borough-Day Level

Borough	Num Rides ( $r_{id}$ )			Price ( $p_{id}$ )			Wage ( $c_{id}$ )		
	q25	median	q75	q25	median	q75	q25	median	q75
Bronx	2181	2334	2654	13.0	13.3	13.7	10.9	11.1	11.4
Brooklyn	3853	4334	5012	16.5	16.8	17.1	13.3	13.6	13.9
Manhattan	6559	7265	7944	18.9	19.8	21.0	15.1	15.9	16.6
Queens	2934	3146	3504	20.3	21.4	22.3	16.6	17.6	18.2

Before introducing the empirical model, we present some data patterns that shed light on connections between our theoretical results and the empirical analysis. Fig. 4 provides evidence suggesting that access is skewed toward denser regions. The first plot on the left shows population density (people per square miles) by borough, while the second plot displays the average number of drop-offs per hour, normalized by region size  $s_i$  as previously defined. Both measures serve as loose proxies for ride demand. The third plot presents the average number of pickups per hour, normalized by  $s_i$ . Perhaps unsurprisingly, as shown in the figure, the rank ordering among the four boroughs is consistent across the first three panels: Manhattan, Brooklyn, Bronx, and Queens. The last panel of Fig. 4 displays the pickup/drop-off ratio for each borough. If every passenger who used Uber to leave a borough also used Uber to return, all these ratios would equal

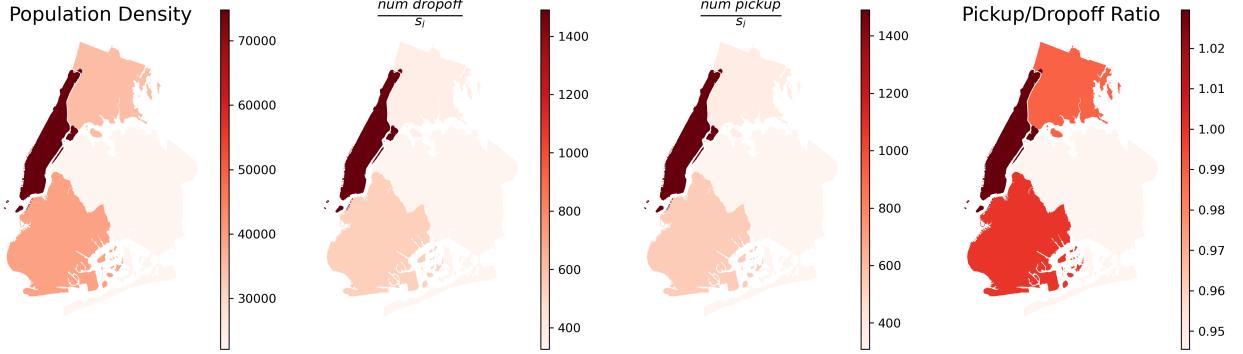
<sup>14</sup>We use gross earnings before expenses rather than net earnings after expenses because our model assumes that the number of drivers is determined by  $c_i$ , the payment from the platform before expenses. Intuitively, for a driver comparing driving on Uber to a minimum-wage job (e.g., at a restaurant), Uber must offer higher gross earnings to account for vehicle maintenance and fuel costs. Parrott and Reich (2018) reports a similar gross earnings of \$22-\$23 for NYC in October 2017 and proposes a minimum gross earnings standard of \$25.76 for NYC drivers.

<sup>15</sup><https://www.census.gov/quickfacts/fact/table/newyorkcountynewyork,richmondcountynewyork,kingscountynewyork,queenscountynewyork,bronxcountynewyork,newyorkcitynewyork/LND110220>

<sup>16</sup>We consider the square root of the borough area as its size based on the following simple hypothetical: if a borough's size doubles in each dimension (quadrupling its area), the average distance between two randomly picked points doubles rather than quadrupling. The notion of measuring the region size in length as opposed to area is also implemented in the micro-foundation for our theory model, as detailed in Appendix C.

1. Interestingly, however, the ratios differ from 1, and follow the same rank ordering as the other measures: Manhattan, Brooklyn, Bronx, and Queens. This suggests that access to rides is easier in denser regions, consistent with the predictions of our theory.

Figure 4: Population Density, Drop-offs, Pickups, and Pickup/Drop-Off Ratios by Borough



Based on the intuition provided by Fig. 4, we now construct a measure that will prove directly useful in our empirical analysis. Using the inter-region rides  $r_{ijd}$ , defined as the average number of rides per hour starting in borough  $i$  and ending in borough  $j$  on a given day  $d$ , we construct the “relative outflow” between  $i$  and  $j$  as follows:

$$RO_{ijd} = \frac{r_{ijd}}{r_{jid}} \quad (6)$$

Table 2 displays all such pair-wise relative outflows from our data, averaged across days. Interestingly, for any two boroughs  $i, j$  where  $i$  is denser than  $j$  in any of the three left panels of Fig. 4, we have  $RO_{ij} > 1$ . Again, this suggests that rides are more accessible in those denser regions, consistent with our theory. Next section further formalizes this idea, clearly sets out the assumptions underneath it, and develops it into an identification/estimation strategy.

Table 2: Average Relative Outflows  $RO_{ij}$  across Days

	region $j$			
	Bronx	Brooklyn	Manhattan	Queens
region $i$				
Bronx	—	0.93	0.93	1.05
Brooklyn	1.07	—	0.91	1.14
Manhattan	1.07	1.1	—	1.25
Queens	0.96	0.87	0.8	—

## 5 Empirical Model

Our theory model characterizes platform, driver and passenger behavior under profit maximizing assumption for platforms and drivers. We now take the model to data to obtain parameter estimates of model primitives and to run counterfactual analyses to provide policy recommendations.

Section 5.1 sets up the empirical calibration problem, specifying observable data and parameters to be calibrated. Section 5.2 presents the empirical model with supply-side EOD. Section 5.3 presents the empirical model with demand-side EOD. Section 5.4 introduces a comprehensive model that incorporates both supply-side and demand-side EOD. Before describing the model, we summarize the notation used for both the theoretical and empirical models in Appendix A.

## 5.1 Connecting the model to data and the role of inter-region rides

This section sets up the empirical calibration problem. We detail a proposition that will prove useful in finding the solution to the problem.

Start by noting that each driver makes  $\frac{1}{W_{id}}$  trips per unit of time, which in our context corresponds to trips per hour. Thus, the number of rides is given by:  $r_{id} = \frac{n_{id}}{W_{id}}$ . Free entry of drivers ensures that drivers' revenue is equivalent to the outside option in equilibrium, implying  $W_{id} = \frac{c_{id}}{c}$ . These relationships imply that the number of drivers can be directly estimated on the data from the number of trips, driver wage, and the outside option:

$$n_{id} = r_{id} \times \frac{c_{id}}{c} \quad (7)$$

Let  $\mathcal{D}$  denote the set of observable data and the variables that can be derived from the data as described above:  $\mathcal{D} = \{p_{id}, c_{id}, r_{id}, n_{id}\}_{i \in I, d \in D} \cup \{r_{ijd}\}_{i, j \in I, d \in D} \cup \{\bar{c}\}$ , where  $I$  denote the set of regions and  $D$  denote the set of days covered in our dataset. The objective of our empirical exercise will be to recover using  $\mathcal{D}$  the following set of parameters:  $\{\bar{\lambda}_{id}\}_{i \in I, d \in D}, \{t_i\}_{i \in I}, \alpha, \beta$ .

In recovering the above parameters, we leverage additional parameterization as well as additional results based on our model. Starting by region sizes  $t_i$ , we assume  $\forall i : t_i = a \times s_i$ , where  $s_i$  is the square root of borough  $i$ 's area (hence observable) and  $a$  is a parameter to be estimated. This parameter could be interpreted as the expected pickup time in a region if the region had one driver per unit of size. We assume  $a$  is homogeneous across regions.<sup>17</sup> This approach reduces the estimation burden for sizes  $t_i$  to only one scalar.

For unobservable potential demand  $\bar{\lambda}_{id}$ , we leverage an assumption that cross-regional demand is balanced:  $\forall i, j \in I, \bar{\lambda}_{ij} = \bar{\lambda}_{ji}$ . The motivation for this assumption is that given our time periods are daily, for every trip there is a “return trip” shortly after. We now show that this assumption, combined with the rest of the structure of our model, creates a tight connection between the unobservable potential demand and the observable ride counts.

**Proposition 6.** *Let  $I$  denote the set of regions served by the platform.  $(n, \rho)$  is a steady-state (not necessarily equilibrium) driver allocation according to Definition 3, under  $(c, p, \bar{\lambda}, t)$ . Suppose potential demand for rides are “balanced”:  $\forall i, j \in I, \bar{\lambda}_{ij} = \bar{\lambda}_{ji}$ . Also suppose the proportion of*

---

<sup>17</sup>In reality,  $a$  could be heterogeneous due to different regional extents of traffic congestion, etc. We abstract from those in this paper. Note that this abstraction likely understates economies of density, given that denser regions tend to have higher traffic.

unfulfilled demand does not depend on destination:  $\forall i, j \in I, \frac{r_{ij}}{\lambda_{ij}} = \frac{r_i}{\lambda_i} = A_i$ .<sup>18</sup> Then we have:

$$\forall i, j \in I, \frac{\frac{r_i}{\lambda_i}}{\frac{r_j}{\lambda_j}} \equiv \frac{A_i}{A_j} = \frac{r_{ij}}{r_{ji}} \equiv RO_{ij}$$

The proof can be found in Appendix F.1. This result shows that the access ratio between two regions is equal to the relative outflow. To interpret the result, if region  $i$  has a lower access than region  $j$ , that is, the supply is more skewed to region  $j$  rather than region  $i$ , then there are more inter-region rides from  $j$  to  $i$  than those from  $i$  to  $j$  following our assumptions. This illustrates how observable relative flows  $RO_{ij}$  help infer the skew among unobservable access levels  $A_i$  and  $A_j$ , which in turn helps calibrate the magnitude of economies of density. We carry out this calibration exercise for both the supply-side and the demand-side models.

Although the assumptions in Proposition 6 are nontrivial, the approach provides meaningful advantages. First, the data requirement is mild: only the numbers of rides rather than vacant cars, search behavior, etc. Second, it can help identify spatial supply-demand mismatch (i.e., skew in access) in passenger transportation markets regardless of whether the matching technology is centralized or decentralized. The measure of potential demand it helps infer (as we do later in the calibration procedure) includes not just the set of passengers who searched for rides by showing up on their app but failed to obtain a ride, but also those who skipped searching for rides in the first place due to anticipating that one would be difficult to find or the rides are too expensive.<sup>19</sup>

With this result at hand, we next turn to empirically calibrating our model.

## 5.2 Supply-Side Economies of Density

We adapt the theory model with supply-side economies of density to the empirical setting here. Next, we describe the procedure to estimate the model parameters with the data. Finally, we show the identification of the model parameters.

In contrast to the theory section, here we do not (need to) assume that the platform optimally sets prices and wages in the empirical model. Instead, we focus on estimating the model parameters from the driver and passenger behavior, and our approach is agnostic to the price and wage setting process. Our focus is on using the estimated model parameters to make recommendations to the

---

<sup>18</sup>A similar proportionality assumption can be found in Bimpikis et al. (2019). See equation (2) on page 748 and the subsequent explanation. One might think an alternative assumption could be that the proportion of fulfilled demand over the actual demand does not depend on the destination, i.e.  $\frac{r_{ij}}{\lambda_{ij}} = \frac{r_i}{\lambda_i}$ . We note that both in our paper and in Bimpikis et al. (2019), these two assumptions are equivalent because of one key underlying assumption: we assume that passengers in any region  $i$  react the same way to prices (and to pickup time in our demand-side model) regardless of their destination. In this way, we have  $\frac{\lambda_{ij}}{\lambda_{ij}} = \frac{\lambda_i}{\lambda_i} = f(p_i, \omega_i)$  for any region  $i, j$ . If one relaxed this assumption and allowed, for instance, those who travel from Manhattan to Bronx to be more/less price-sensitive than those who travel from Manhattan to Brooklyn, then  $r_{ij}/\lambda_{ij} = r_i/\lambda_i$  would no longer be equivalent to  $r_{ij}/\lambda_{ij} = r_i/\lambda_i$ .

<sup>19</sup>To illustrate, suppose passengers consistently use rideshare to go to region  $i$  but never open their apps to find a ride back out of region  $i$  because they have learned that supply is scarce in that region. Even hard-to-obtain data on app sessions cannot detect this phenomenon and would, hence, underestimate the low access to rides in region  $i$ , but a method based on relative ride flows can.

platform in our counterfactual analysis.

Recall from the theory section 3.2, for each region  $i$  and each day  $d$ , given the market primitives  $(\bar{\lambda}_{id}, t_i, \bar{c}, \alpha)$  and  $(p_{id}, c_{id})$  chosen by the platform, the equilibrium number of drivers is given by

$$\frac{c_{id}}{W_{id}(n_{id})} = \bar{c}$$

Following the theoretical model, the wait time is given by  $W_{id} = \frac{n_{id}}{\bar{\lambda}_{id}(1-\alpha p_{id})} + \frac{t_i}{n_{id}}$ .<sup>20</sup>

We had operationalized the region size parameter  $t$  as  $t_i = as_i$ , where  $s_i$  is the square root of the region (borough) size, and is therefore observable. Thus, having  $n_{id}$  estimated from eq. (7) and observable platform choices  $p_{id}$  and  $c_{id}$ , the driver free-entry equilibrium ensures that the wait time obtained equates the drivers' ridesharing wage to the outside option. This leads the following estimating equation:

$$\frac{n_{id}}{\bar{\lambda}_{id}(1 - \alpha p_{id})} + \frac{as_i}{n_{id}} = \frac{c_{id}}{\bar{c}} \quad (8)$$

The parameters to be estimated in the supply-side model are, therefore,  $(\alpha, a, \{\bar{\lambda}_{id}\}_{i \in I, d \in D})$ .

## Calibration

The model calibration follows two steps. First, for any candidate pair  $(\alpha, a)$ , we use the model structure and observable data to recover the potential demand  $\bar{\lambda}_{id}$  for each region  $i$  and day  $d$ . Next, we use these recovered potential demand values to simulate relative outflows between pairs of regions, which are inherently a function of the initial  $(\alpha, a)$ . We then match these simulated relative outflows to the observed data by minimizing the sum of squared differences, yielding estimates  $(\hat{\alpha}, \hat{a})$  (McFadden, 1989). Formally, our steps are as follows:

**Step 1** We obtain the number of drivers from the data following eq. (7). Using the number of drivers, eq. (8) allows us to express  $\bar{\lambda}_{id}$  in terms of data  $(p_{id}, c_{id}, s_i)$  and now  $n_{id}$  and parameters  $(a, \alpha)$  as follows:

$$\bar{\lambda}_{id}(a, \alpha; \mathcal{D}) = \frac{n_{id}}{(1 - \alpha p_{id})(\frac{c_{id}}{\bar{c}} - \frac{as_i}{n_{id}})}, \quad (9)$$

**Step 2** Given the observables and following step 1, we express the relative outflow given by the empirical model between each pair of regions  $i, j$  on each day  $d$  as a function of the parameters  $(a, \alpha)$ :

$$RO_{ijd}^{model} = \frac{A_{id}}{A_{jd}} = \frac{r_i / \bar{\lambda}_{id}(a, \alpha; \mathcal{D})}{r_j / \bar{\lambda}_{jd}(a, \alpha; \mathcal{D})},$$

where the first equality is given by Proposition 6, and the second equality is given by the definition of access as in eq. (3). We estimate  $(a, \alpha)$  by matching the moments of  $RO_{ij}$  for every pair of regions  $i, j$ , where we take the average difference in  $RO_{ij}$  across  $d \in D$  between the model simulation and the data. Specifically, for each pair of regions  $i, j$ , denote

---

<sup>20</sup>We assume that the trip time is 0 as in our theory model.

$d_{ij}(a, \alpha; \mathcal{D}) = |\frac{1}{|\mathcal{D}|} \sum_d RO_{ijd}^{model}(a, \alpha; \mathcal{D}) - \frac{1}{|\mathcal{D}|} \sum_d RO_{ijd}^{data}(a, \alpha; \mathcal{D})|$  and let  $d(\alpha, a; \mathcal{D})$  be the vector consisting of  $d_{ij}$ . To obtain the estimates  $(\hat{\alpha}, \hat{a})$ , we minimize

$$\min_{\alpha, a} d(\alpha, a; \mathcal{D})' d(\alpha, a; \mathcal{D}), \quad (10)$$

The estimates of  $\bar{\lambda}_{id}$  are then obtained from  $\bar{\lambda}_{id}(a, \alpha; \mathcal{D})$  following eq. (9).

### 5.2.1 Identification

We informally discuss the identification of model parameters here, and relegate the formal statement and proof of the identification to Appendix F.2. Below, we outline the intuition behind this process.

Using the number of drivers obtained from eq. (7), eq. (8) allows us to express  $\bar{\lambda}_{id}$  in terms of data and parameters  $(a, \alpha)$ . Next, observe that each  $\bar{\lambda}_{id}$  can be written as an explicit function of  $a, \alpha$  and data as in eq. (9). Therefore, the identification problem involves only the parameters  $\alpha$ , and  $a$  using variation in the RO moments. We consider the relative outflow across regions for a single day  $d$ . For notational brevity, we drop the day subscript  $d$ . The model relative outflow can be written as

$$RO_{ij}^{model} = \frac{A_i}{A_j} = \frac{r_i/\bar{\lambda}_i}{r_j/\bar{\lambda}_j} = \frac{\frac{n_i}{c_i}/\bar{\lambda}_i}{\frac{n_j}{c_j}/\bar{\lambda}_j} = \frac{(1 - \alpha p_i) \left(1 - a \frac{s_i \bar{c}}{n_i c_i}\right)}{(1 - \alpha p_j) \left(1 - a \frac{s_j \bar{c}}{n_j c_j}\right)} \quad (11)$$

The third equality holds given the assumption of driver equilibrium distribution. The fourth equality is obtained if we substitute  $\bar{\lambda}_i$  and  $\bar{\lambda}_j$  with  $a, \alpha$  and observable data  $\mathcal{D}$  following eq. (9). Note that in this equation, everything is data except for  $a$  and  $\alpha$ .

The formal proof of how matching the moments in eq. (10) identifies both  $a$  and  $\alpha$  is relegated to Appendix F.2. To see the intuition, first let us assume  $\alpha = 0$ . In this case, roughly,  $a$  is identified by the variation in relative outflows: the larger the magnitude of  $a$ , the larger the imbalance of flows between denser and less dense regions.<sup>21</sup> In addition, variation in wage levels over time and regions also contributes to the identification.

Price sensitivity  $\alpha$  would be identified in a similar fashion if we assumed  $a = 0$ : the larger the magnitude of  $\alpha$ , the larger the imbalance of flows between more and less expensive regions.<sup>22</sup> Now, note that if there is sufficient independence between regional price variation and regional density variation, both  $a$  and  $\alpha$  are identified.<sup>23</sup>

---

<sup>21</sup>More specifically,  $RO_{ij}$  being close to 1 is consistent with low values of  $a$ . As  $a$  increases,  $RO_{ij}$  would need to diverge from 1. For example, if  $\frac{s_i \bar{c}}{n_i c_i} > \frac{s_j \bar{c}}{n_j c_j}$ , then  $RO_{ij}$  is less than 1 and it decreases as  $a$  increases.

<sup>22</sup>In principle, the price variation across region is prone to endogeneity, which might bias the results. One possible solution to this problem would be to calibrate the parameter  $\alpha$  in a way that would lead to the a similar price elasticity of demand to what is found in other research that leverages exogenous variation for the same market (such as Rosaia (2020)), instead of directly recovering it from our data. We did not do so, however, given that the  $\alpha$  we found indeed implies price elasticity levels that are close to Rosaia (2020) (See Table 4 for the results).

<sup>23</sup>Broadly speaking, sufficient independence requires that the comparison of  $p$  between two regions cannot inform how  $\frac{s \bar{c}}{n c}$  compares between the two regions, and vice versa. Specifically, there exists a pair of regions where both  $p$  and  $\frac{s \bar{c}}{n c}$  are high for one of the regions; and there also exists another pair of regions where one region has higher  $p$  and another region has higher  $\frac{s \bar{c}}{n c}$ . See the proof in Appendix F.2 for more details.

### 5.3 Demand-Side Economies of Density

In this section, we adapt the theory model with demand-side economies of density to the empirical setting. The overall strategy is analogous to the case of supply-side EOD, so we skip details when they are similar. We describe the estimation procedure and identification for this model in Appendix D.

Again, we do not assume that the platform optimally sets prices and wages and focus on estimating the model parameters from the driver and passenger behavior. Following the theoretical model with demand-side economies of density, the wait time is given by  $W_{id} = \frac{n_{id}}{\bar{\lambda}_{id}(1-\alpha p_{id})\left(1-\beta' \frac{t_i}{n_{id}}\right)}$ . Recall from our discussion in the theory section 3.3 that to model demand-side-*only* EOD, we need to assume that  $t_i \equiv a \times s_i$  is vanishingly small but  $\beta'$  is so large as to make  $\beta s_i \equiv \beta' \times a \times s_i$  of non-trivial magnitude. Thus, in equilibrium, we will have:

$$\frac{c_{id}}{W_{id}(n_{id})} = \bar{c} \iff \frac{n_{id}}{\bar{\lambda}_{id}(1-\alpha p_{id})\left(1-\frac{\beta s_i}{n_{id}}\right)} = \frac{c_{id}}{\bar{c}} \quad (12)$$

The parameter  $\beta$  can be interpreted as the reduced-form aggregation of (i) the model parameter of pickup time as a function of region size and the number of drivers (ii) the sensitivity of ride demand to pickup time. Therefore,  $\beta$  captures the demand-side EOD. In this model, the parameters to be estimated are  $(\alpha, \beta, \{\bar{\lambda}_{id}\}_{i \in I, d \in D})$ .

### 5.4 The Comprehensive Model

In taking the model to data, we construct a comprehensive model here that includes both supply-side and demand-side economies of density that we have separately examined above. In the comprehensive model, the passenger pickup time affects both the demand arrival rate and drivers' supply decisions, where the total wait time per ride is given by

$$W_{id}(n_{id}; \bar{\lambda}_{id}, \alpha, \beta, a) = \frac{n_{id}}{\bar{\lambda}_{id}(1-\alpha p_{id})\left(1-\frac{\beta s_i}{n_{id}}\right)} + \frac{a s_i}{n_{id}} \quad (13)$$

The distribution of drivers on a given day  $d$  is still obtained as an equilibrium outcome, where the ratio of wage per ride to wait time equals the value of the outside option:

$$\forall i, \frac{c_i}{W_{id}(n_{id}; \bar{\lambda}_{id}, \alpha, \beta, a)} = \bar{c} \quad (14)$$

As before, the number of rides is given by  $r_{id} = \frac{n_{id}}{W_{id}}$ .

The parameters to be estimated are  $(a, \alpha, \beta, \{\bar{\lambda}_{id}\}_{i \in I, d \in D})$ . This general model encompasses, as special cases, the two models we have examined so far: the supply-side-only EOD ( $\beta = 0$ ) and the demand-side-only EOD ( $a = 0$ ). As discussed, separately identifying  $a$  and  $\beta$  is not feasible given our data. As demonstrated earlier, each parameter can be identified if we assume the other is zero. Similarly,  $a$  and  $\beta$  would be identified if we assume that the ratio between them is known.

This effectively translates to the idea that the relative contributions of the supply-side and the demand-side EOD are known.

We can use this idea to test for robustness across a wide range of relative contributions of these effects. Specifically, we calibrate the model using the following ratios:  $\frac{a}{\beta} \in \{10, 1, 0.1\}$ . A larger ratio implies that supply-side effects are stronger, while a smaller ratio implies that demand-side effects have a greater contribution. Observe that a model estimated with each of these assumptions on  $\frac{a}{\beta}$  is in nature “arbitrary” and uninformative on its own, since we cannot test this assumption. However, the analysis across this wide range of  $\frac{a}{\beta}$  is useful since it helps to show that our results are robust across a wide range of possible ratios of the relative importance between supply-side and demand-side EOD, as opposed to just the two extreme cases.<sup>24</sup>

## 5.5 Empirical Results

We present estimates from the models with supply-side EOD, demand-side EOD, and the comprehensive model with both sources of EOD. Table 3 presents the estimates of the parameters from the supply-side and the demand-side model. The parameter estimates are qualitatively similar. Table 4 shows the price elasticities implied by the model estimates. As a benchmark, Rosaia (2020) finds a broadly consistent price elasticity ranging from 0.55 to 1.09.<sup>25</sup> Cohen et al. (2016) finds an overall price elasticity of 0.61 for the NYC.

Table 3: Parameter Estimates from the Supply- and Demand-Side Models

	Supply-Side Model	Demand-Side Model
$\alpha$	0.0201 (33.36)	0.0138 (21.17)
$a$	23.27 (29.82)	N/A
$\beta$	N/A	42.64 (26.88)
Average $\bar{\lambda}_{id}$ by Regions		
Bronx	4674	3943
Brooklyn	7824	6791
Manhattan	12650	10535
Queens	6565	5629

Note: We report the point estimates (with t-statistics in parentheses) for the parameters ( $\alpha, a, \beta$ ). Since our model estimates  $\bar{\lambda}_{id}$  for each day and each region, we therefore only present the average of the point estimates of  $\bar{\lambda}_{id}$  across days for each region.

Table 5 displays the average potential demand densities and accesses across days in each region, as estimated by each model. The results reveal significant variation in the demand density across regions, with Manhattan exhibiting more than four times the density of Queens, according to both the models. This spatial disparity highlights the importance of characterizing and estimating these

<sup>24</sup>In addition to  $\frac{a}{\beta} \in \{10, 1, 0.1\}$ , we also ran the analysis with  $\frac{a}{\beta} \in \{0.01, 100\}$ . The results were similarly robust.

<sup>25</sup>Rosaia (2020) is an order version of Rosaia (2023). It reports the price elasticities implied by the model estimates.

Table 4: Price Elasticities by Regions

Region	Supply-Side Model	Demand-Side Model
Bronx	0.37	0.34
Brooklyn	0.51	0.36
Manhattan	0.67	0.40
Queens	0.75	0.55
All Regions	0.60	0.41

Note: The table shows the average demand elasticities across days with respect to prices given by supply-side and demand-side models separately. The averages are weighted by the number of rides,  $r_{id}$ .

regional differences.

Furthermore, the access levels follow an ordering that is consistent with the demand density. Specifically, Queens, the least dense region, is found to have the lowest access and Manhattan, the most dense region, the highest access. These results remain robust across varying model specifications that allow for different relative levels of supply-side and demand-side EOD, as detailed in Appendix G, where we present the demand densities and accesses from the alternative models. With model calibration results at hand, we next turn to counterfactual analysis.

Table 5: Model Outcomes (Regions Ordered by Demand Density)

Model Spec.	Supply-Side Model	Demand-Side Model
<i>Demand Density (<math>\frac{\bar{\lambda}_i}{s_i}</math>)</i>		
Queens	630	540
Bronx	719	607
Brooklyn	939	815
Manhattan	2655	2211
<i>Access (<math>\frac{r_i}{\lambda_i}</math>)</i>		
Queens	0.49	0.57
Bronx	0.51	0.61
Brooklyn	0.57	0.66
Manhattan	0.58	0.69

Note: The results presented correspond to averages across days over the data time periods.

## 6 Counterfactual Analysis

We next simulate several counterfactual scenarios using the estimated model parameters. Our primary objective is to empirically quantify the key theoretical insights presented in Section 3. First, we simulate the platform’s optimal price and wage policy in Section 6.1 to assess the extent to which economies of density skews access away from less dense regions, and to what degree the platform might counter such a skew. We demonstrate that the platform mitigates but does not eliminate the access skew by comparing the access skew under the optimal heterogeneous and

homogeneous pricing strategy. Second, we simulate the platform’s optimal policy with the demand levels in each region scaled by a uniform factor, as described in Section 6.2, to quantify how much the access skew becomes more pronounced for smaller platforms. Finally, in Section 6.3, we go beyond quantifying the theoretical insights and explore the region-specific prices/wages necessary to achieve equalized access across regions.

Our counterfactual analysis reveals findings that are robust across different models, regardless of whether and to what extent economies density originates from the supply-side, the demand side, or both. Note that we do not have theory results in the case of mixed supply-and-demand EOD, but we can obtain counterfactual findings for this case too. The counterfactual demonstrates that our insights are consistent across all models empirically. To avoid redundancy, we provide the counterfactual results only for the supply-side model here. Appendix G has the results for all models: the supply-side model, the demand-side model, and the model with both supply-side and demand-side EOD.

## 6.1 Platform Optimal Strategy

Recall that our estimation did no rely on any assumptions about the platform’s pricing policy. In this section, we first simulate the platform’s optimal prices and wages to quantify the idea that the platform might set higher prices and wages but accepts lower margins in less dense regions. Further, we examine the platform’s responses when it is restricted – either by regulation or company policy – to implement a uniform price and wage across all regions. Comparing these scenarios, we find that while the platform does not fully eliminate spatial access skew, its optimal strategy is to mitigate access skew using flexibility in setting prices and wages across regions.

Based on our model estimates, we first simulate the platform’s optimal prices and wages across each of the regions for each day over the four-month span of data. Fig. 5a illustrates the optimal prices and wages for each region (averaged across days), with regions arranged in descending order by demand density from left to right. The platform’s optimal policy pays drivers approximately 77% more per ride in the least dense region (Queens) compared to the most dense region (Manhattan). However, the platform does not pass on more than 25% of this wage difference to passengers. Overall, our simulation recommends a higher price relative to what Uber charged during the data period. This recommendation aligns with Uber’s more recent pricing adjustments that have been implemented after the period of our data.<sup>26</sup>

Next, we examine access levels. Fig. 5b displays the access levels across regions (averaged across days) under the optimal heterogeneous and homogeneous price and wage policies.<sup>27</sup> Two key patterns emerge. First, lower-density regions exhibit lower access levels. Under optimal heterogeneous prices and wages, access in the most dense region (Manhattan) is higher by 25% compared to the least dense region (Queens).

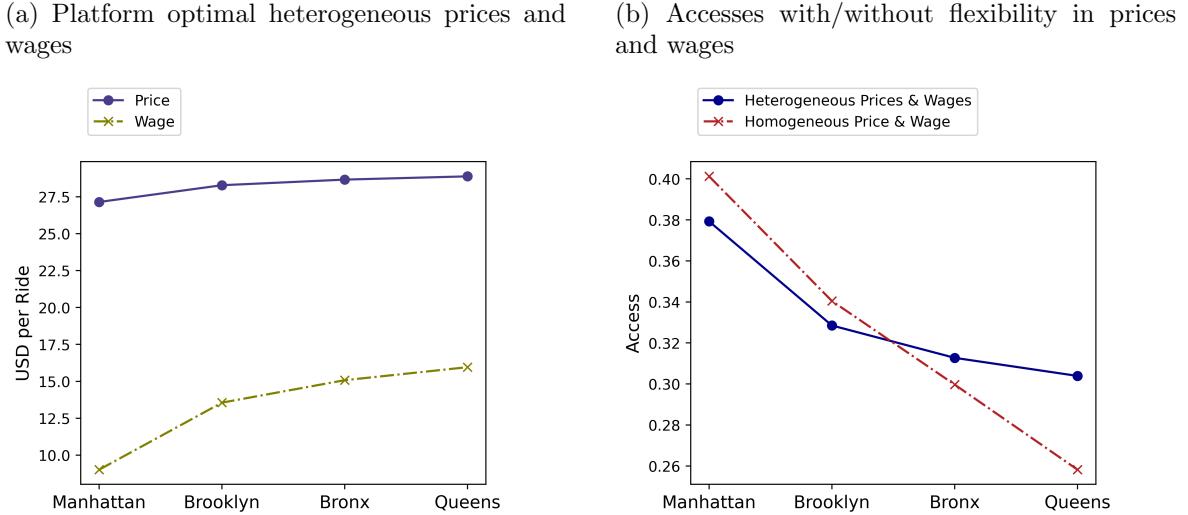
---

<sup>26</sup>See Appendix I for more details.

<sup>27</sup>Note that compared to accesses level calibrated from the model shown in Table 5, the accesses here are at a relatively lower level. This is because our model predicts prices that are higher than the observed levels.

Second, compared to the scenario where the platform is restricted to set uniform price and wage across regions, the platform mitigates but does not fully eliminate access inequality across regions with pricing flexibility. Specifically, as Fig. 5b illustrates, access in the most dense region is higher by 55% than the least dense region under homogeneous pricing, but only 25% higher under heterogeneous pricing. This aligns with our interpretation of Proposition 3 and Proposition 4, which suggests that the platform optimally charges smaller margins in sparser regions under full pricing flexibility (effectively “subsidizing” sparser regions) in order to build economies of density and mitigate spatial access skew.

Figure 5



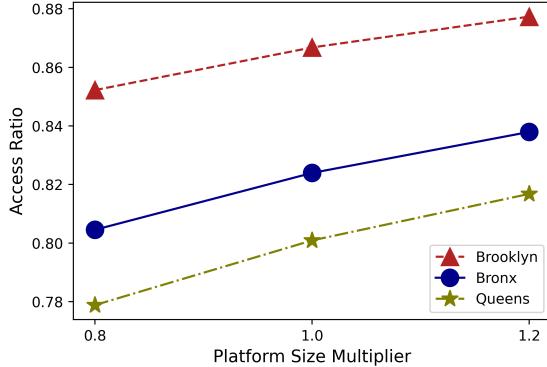
Note: The regions (boroughs) are arranged in descending order by demand density from left to right.

## 6.2 Impact of Platform Size

There have been regulations that cap the number of drivers on ridesharing platforms, effectively reducing platform size (CBS News, 2018). To illustrate how the degree of access skew changes as the market thickens, we scale the demand levels for each region by a uniform factor to simulate a larger or smaller platform. This allows us to observe how access skew is either mitigated or intensified. We use the average potential demand ( $\bar{\lambda}_{id}$ ) across days as the base demand levels for each region. We then scale these demand levels by a uniform platform size multiplier, with values set to 0.8 and 1.2. Following this adjustment, we calculate the optimal prices and wages as well as the associated access levels in equilibrium. Under the supply-side EOD model, if Uber’s size were to shrink by 20%, then the access of the most dense region is 28% higher than the least dense region. On the other hand, if Uber’s potential demand grows by 20% of the estimated size, the access in Manhattan is higher by 22% than Queens. Consistently, as depicted in Fig. 6, the access ratios of the three less dense regions (Brooklyn, Bronx and Queens) relative to Manhattan approach 1 as the platform expands. This indicates that the access skew is mitigated as the platform grows

larger.

Figure 6: Access Ratio (relative to Manhattan) as market thickens with supply-side EOD model.



### 6.3 Equalizing Accesses

Policy makers are sensitive to unequal access to important services like transportation. We examine the extreme case where the policy maker might want to equalize access across all regions to provide insight into what such a policy would require. Specifically, we evaluate the extent to which Uber would need to adjust its regional prices/wages in order to achieve uniform access across regions, set to the average access level across regions, which is closest to that of Brooklyn.

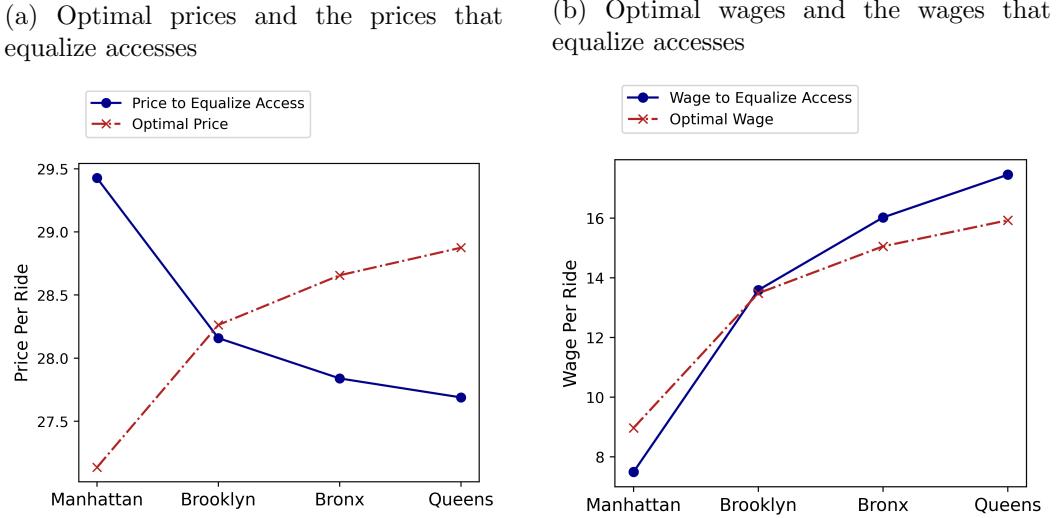
We set the potential demands of each borough at the estimated ones (averaged across days). We then fix either wages ( $c^*$ ) or prices ( $p^*$ ) at the optimal level as determined when the platform jointly optimizes  $(p^*, c^*)$ . We then vary the other lever – prices or wages accordingly.

First, we fix the wage at  $c^*$  and then calculate the regional price levels  $p_i$  needed to equalize access across all regions (at the average access level under the full equilibrium). We find that, as shown in Fig. 7a, to equalize access, the upward trend in prices (where prices are higher in sparser regions) is reversed, resulting in a downward trend. Specifically, the price in the most dense region increases by 8.4%, while the price in the least dense region decreases by 4.1%.

Next, we fix prices at  $p^*$  and adjust regional wages to set all access levels to the global average. The upward trend in wages as demand density decreases becomes more pronounced, as illustrated in Fig. 7b. To achieve equal access, the wage in the most dense region needs to be reduced by 16.5%, while the wage in the least dense region must increase by 9.6%. This adjustment further amplifies the wage disparity across regions.

These findings align with the idea that, to build economies of density, the platform increases wages in less dense regions, but does not pass these costs fully onto passengers. Furthermore, under optimal prices and wages, the margin in the least dense region is 29% lower than in the most dense region, a percentage that increases to 43% and 42% when accesses are equalized using price and wage levers respectively. This implies that the sparser regions are effectively subsidized to an even greater extent.

Figure 7



Note: Consider the prices and wages denoted as  $(p^*, c^*)$ , determined under the optimal heterogeneous pricing scheme described in Section 6.1. The left panel illustrates a comparison between  $p^*$  and the prices required to achieve equal access when the wages are fixed at  $c^*$ . The right panel compares  $c^*$  with the wages necessary to ensure equal access when prices are fixed at  $p^*$ . The regions (boroughs) are arranged in descending order by demand density from left to right.

## Robustness of Findings

We carry out a battery of additional robustness checks to evaluate the validity of the findings in the paper. First, we adjust for and incorporate trip time into the empirical model. Specifically, we construct a price measure adjusted for trip distance and time and incorporate an additional average trip time in the model. We re-estimate the extended model with this new price measure. See Appendix H for more details on how we implement this. Second, we allow for region-dependent parameters, such as price sensitivity, in both theory and empirical models. Third, we conduct empirical analysis using finer grid partitions of each NYC boroughs. Finally, we examine alternative functional forms for modeling pickup time and customer response to price and pickup time. Our main insights continue to hold across all of these specifications.<sup>28</sup>

## 7 Conclusion

This paper examines the effects of economies of density and market thickness on spatial distribution of access to service. We focus on ridesharing and analyze this market both theoretically and empirically. Theoretically, we show that in equilibrium, relative to lower density regions, regions with higher densities of potential demand get more supply—even after normalizing by their higher demand. We show that this “spatial access skew” is more intense for smaller platforms. Finally, we show that a ridesharing platform’s optimal pricing strategy would involve mitigating, but not

<sup>28</sup>The complete set of results for these robustness checks are available from the authors.

fully eliminating, the spatial mismatch between supply and demand. All of these results are robust to whether the source of economies of density is the supply-side or the demand-side.

Empirically, we calibrate our model using ride-level data on Uber from New York City. A key aspect of our calibration exercise is the identification of relative access to service across regions, achieved by leveraging the “relative outflows” of rides between pairs of regions. The core idea behind our strategy for identifying the extent of economies of density is broadly applicable beyond our specific context to all passenger-transportation markets, whether the matching technology is centralized (e.g., ridesharing) or decentralized (e.g., taxicabs). Using our calibrated model, we address several important questions, including platform optimal prices and wages across regions, the impact of constraining the platform to set spatially homogeneous price and wage, the role of platform size (i.e., market thickness), and the extent of price/wage adjustment needed to homogenize access across regions.

The analysis in this paper can be extended in multiple directions, especially if additional driver-level and passenger-level data are available. First, extending the model to markets with decentralized matching, such as the taxicab market, would provide useful insights on the relative effectiveness of policies in that market compared to ridesharing. Second, incorporating a trip-level structural model could help explore how trip-level dynamics contribute to access differences, allowing for heterogeneity in trip characteristics. Such an extension would also facilitate the empirical disentanglement of supply- and demand-side economies of density. This would also allow us to better examine competitive dynamics between platforms (Rosaia, 2023). Third, research could examine the differing incentives of social planners and platforms in terms of access differences across regions, and the conditions under which they converge and diverge. Fourth, a useful next step would be to examine the tradeoff between platform size, which we find to have a positive effect here, with the effect of competition, in considering whether platforms should be allowed to coordinate or even merge. Finally, understanding the value of other policy instruments like regulating the total number of drivers in a region or congestion pricing would be of broad interest in other spatial markets beyond ridesharing.

## References

- Afèche, P., Liu, Z., and Maglaras, C. (2023). Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Manufacturing & Service Operations Management*, 25(5):1890–1908.
- Akbarpour, M., Li, S., and Gharan, S. O. (2020). Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815.
- Ashlagi, I., Burq, M., Jaillet, P., and Manshadi, V. (2019). On matching and thickness in heterogeneous dynamic markets. *Operations Research*.

- Banerjee, S., Kanoria, Y., and Qian, P. (2018). The value of state dependent control in ridesharing systems. *arXiv preprint arXiv:1803.04959*.
- Besbes, O., Castro, F., and Lobel, I. (2021). Surge pricing and its spatial supply response. *Management Science*, 67(3):1350–1367.
- Bimpikis, K., Candogan, O., and Saban, D. (2019). Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769.
- Brancaccio, G., Kalouptsidi, M., Papageorgiou, T., and Rosaia, N. (2023). Search frictions and efficiency in decentralized transport markets. *The Quarterly Journal of Economics*, 138(4):2451–2503.
- Buchholz, N. (2022). Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *The Review of Economic Studies*, 89(2):556–591.
- Castillo, J. C., Knoepfle, D., and Weyl, E. G. (2024). Matching and pricing in ride hailing: Wild goose chases and how to solve them. *Management Science*.
- CBS News (2018). New york city council puts cap on uber, other ride-hailing services.
- Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016). Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.
- Cramer, J. and Krueger, A. B. (2016). Disruptive change in the taxi business: The case of uber. *American Economic Review*, 106(5):177–82.
- Diao, M., Kong, H., and Zhao, J. (2021). Impacts of transportation network companies on urban mobility. *Nature Sustainability*, 4(6):494–500.
- Fortune Business Insights (2021). Ride Sharing Market Size, Share & COVID-19 Impact Analysis. Market research report, Fortune Business Insights.
- Frechette, G. R., Lizzeri, A., and Salz, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, 109(8):2954–2992.
- Hall, J. V. and Krueger, A. B. (2018). An analysis of the labor market for uber's driver-partners in the united states. *Ilr Review*, 71(3):705–732.
- Jin, S. T., Kong, H., and Sui, D. Z. (2019). Uber, public transit, and urban transportation equity: A case study in new york city. *The Professional Geographer*, 71(2):315–330.
- Lam, C. T. and Liu, M. (2017). Demand and consumer surplus in the on-demand economy: the case of ride sharing. *Social Science Electronic Publishing*, 17(8):376–388.

- McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026.
- Nikzad, A. (2018). Thickness and competition in ride-sharing markets. Technical report, Working paper, Stanford University.
- Parrott, J. A. and Reich, M. (2018). An earnings standard for new york city’s app-based drivers: Economic analysis and policy assessment.
- Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature*, 39(2):390–431.
- Rosaia, N. (2020). Competing platforms and transport equilibrium: Evidence from new york city.
- Rosaia, N. (2023). Competing platforms and transport equilibrium.
- Shapiro, M. H. (2018). Density of demand and the benefit of uber.

## Appendices

This section provides ten appendices. Appendix A summarizes the notation used in the paper. Appendix B provides anecdotal evidence for the relevance of our model. Appendix C provides a micro-foundation based on a circular city model for the total wait time formula that we use throughout the paper. Appendix D provides calibration and identification details of the empirical model with demand-side EOD (see main paper for supply-side EOD). Appendix E gives proofs of the propositions for the theory model. Appendix F presents proofs of identification results related to the empirical model and showing sufficiency of relative outflow in obtaining potential demand. Appendix G presents the counterfactual results of all variations of the empirical models. Appendix H presents the empirical analysis incorporating trip time. Appendix I shows the price trend after 2019. Appendix J compares the data used in our analysis to that in Rosaia (2023).

### A Table of Notations

Table 6 summarizes the notations we use in both the theoretical and empirical models.

### B Anecdotal Evidence of Driver Factors

This appendix points to a list of anecdotal pieces of evidence (by no means exhaustive) from online rideshare forums on how drivers complain about Lyft's far pickups in suburbs and how they recommend responding to it. The explanations in brackets within the quotations are from us.

- **From the online forum “Uber People,”<sup>29</sup> a thread in the Chicago section:** The title of the thread is “To those who drive Lyft in the suburbs.” The thread was started on Dec 19 2016. The first post says “Are the ride requests you get on Lyft always seem to be far away from your location? Seems like they are always 5 miles or more for the pickup location. I got one for 12 miles last night. I drive in the Schaumburg/Palatine area [two northwestern suburbs of Chicago about 30mi away from downtown].”
- **From the same thread:** “iDrive primarily in Palatine. about two out of every five ride requests are for more than 10 minutes away. I ignore those”.
- **From the same thread:** “I was a victim of that once. Never again I take a ping more than 10 minutes away in the burbs”.
- **From the same thread:** “Yesterday was my 1st day on Lyft. Was visiting in Homer Glen [a village about 30mi southwest of downtown Chicago] & decided to try Lyft for the first time. First ping was 18 minutes away. Dang, I could make it 1/2 way downtown in that time! I ignored the ride request. 2nd ping was also 18 minutes away. Lyft app complained my acceptance rate is too low. I ignored the 2nd ping & went off-line.”

---

<sup>29</sup>In spite of what the name suggests, this is a general ridesharing forum, not exclusively about Uber.

Table 6: Table of Notations

Symbol	Category		Meaning
	Theory Model	Empirical Model	
$i$	Index	Index	Region(Borough) index
$j$	Index	Index	Region(Borough) index
$d$	NA	Index	Day index
$t_i$	Market Primitive	Model Parameter	Size of region $i$
$a$	Market Primitive	Model Parameter	Ratio of $t_i$ to $s_i$
$\alpha$	Market Primitive	Model Parameter	Passenger's price sensitivity
$\beta$	Market Primitive	Model Parameter	Passenger's sensitivity to pickup time
$\bar{\lambda}_i, \bar{\lambda}_{id}$	Market Primitive	Model Parameter	Potential demand of rides per hour of region $i$ (on day $d$ )
$s_i$	NA	Data	Square root of area of borough $i$
$\bar{c}$	Market Primitive	Data	Reservation level of wage
$p_i, p_{id}$	Decision Variable	Data/Decision Variable	Price per ride of region $i$ (on day $d$ )
$c_i, c_{id}$	Decision Variable	Data/Decision Variable	Wage per ride of region $i$ (on day $d$ )
$n_i, n_{id}$	Outcome Variable	Data/Outcome Variable	Number of drivers in region $i$ (on day $d$ )
$W_i, W_{id}$	Outcome Variable	Outcome Variable	Wait time (pickup time plus idle time) per ride in region $i$ (on day $d$ )
$r_i, r_{id}$	Outcome Variable	Data/Outcome Variable	Number of rides per hour originating from region $i$ (on day $d$ ) (destination could be region $i$ or other regions)
$r_{ijd}$	NA	Data/Outcome Variable	Number of rides picked up from region $i$ and dropped off in region $j$ per hour on day $d$
$RO_{ijd}$	NA	Data/Outcome Variable	Relative outflow between region $i$ and region $j$ on day $d$
$\mathcal{D}$	NA	Data	Set of observable data and the variables that can be derived directly from data
$A_i, A_{id}$	Outcome Variable	Outcome Variable	Access of region $i$ (on day $d$ )

- **From the same forum, a thread titled “First 3days of Lyft”:** “If your area is spread out...and you have to take those > 10 minute requests, well...I might look for another job.”
- **From the same thread:** “Yeah, another (mostly) Lyft-specific problem, especially when working in the suburbs, is you sometimes (fairly frequently, actually) receive trip requests that are not close to your current location. I’ve received requests from passengers 20 miles away.”
- **From Chicago Tribune article titled “Lyft takes on Uber in suburbs”:** Jean-Paul Biondi, Chicago marketing lead for Lyft is quoted to explain the reason for Lyft’s planned expansion into suburbs as follows “The main reason is we saw a lot of dropoffs in those

areas, but people couldn't get picked up in those areas." Which is in line with our reasoning that small relative-outflow is a sign of potential demand which does not get served due to under-supply.

- **From the rideshare website “Become a Rideshare Driver”:** It says successful Lyft drivers use the following strategy:
  - “The drivers usually run the Lyft app exclusively when they are in the busy downtown or city areas.”
  - “Usually in the suburbs, Uber is busier than Lyft, and in such areas, the drivers run both the Uber and Lyft apps.”

## C Micro Foundation for the Total Wait Time Formula

This section provides a simple micro-foundation for the total wait time formula eq. (1) using a circular city model ala Salop. The circular model of regions is illustrated in Figure 8. The  $n_i$  drivers (in the figure,  $n_i = 5$ ) are placed on equidistant locations on the circumference. <sup>30</sup> Drivers are matched to arriving passengers via a centralized matching system. Each driver's “range” or “catchment area” will be the arc consisting of all the points on the circle that are closer to that driver than they are to any other driver in region  $i$ . Given that the total arrival rate in the region is  $\lambda_i$ , the arrival rate in each driver's catchment area is  $\frac{\lambda_i}{n_i}$ . Each driver picks up the first passenger that arrives within that driver's catchment area. In practice, ridesharing platforms implement a similar matching rule (Frechette et al. (2019) use a similar approach to model a centralized matching market).

Suppose the time it takes a driver to travel a full circumference to pick up a passenger is  $t'$ . The platform allocates an arriving customer to the closest driver. Because drivers are situated at equidistant points on the circumference of the region, their catchment areas include half the distance to their nearest neighbors on both sides. The idle time expected for a customer to arrive in the driver's area is  $\frac{n_i}{\lambda_i}$ . The distance between drivers is  $\frac{l}{n_i}$  where  $l$  is the circumference. Since consumer location is uniform, the distance a consumer will be from the driver along the arc is distributed  $d \sim U[0, \frac{l}{2n_i}]$ , implying that the expected distance is  $E[d] = \frac{l}{4n_i}$ . Thus, the expected pickup time is  $\frac{t'}{4n_i}$ .

Based on the above, the expected total wait time  $W_i(n_i)$  defined from the driver's perspective as:

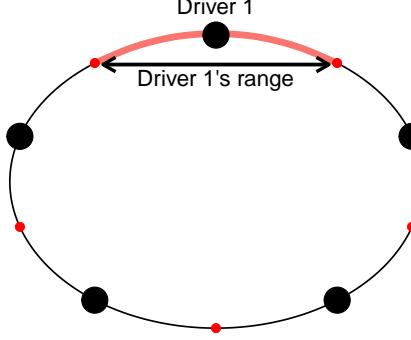
$$\underbrace{W_i(n_i)}_{\text{Total Wait Time}} = \underbrace{\frac{n_i}{\lambda_i}}_{\text{Idle Time}} + \underbrace{\frac{t'}{4n_i}}_{\text{Pickup Time}} = \left( \frac{n_i}{\lambda_i} + \frac{t_i}{n_i} \right) \quad (15)$$

where  $t = \frac{t'}{4}$ . This is exactly the same as eq. (1) except the notation on  $\lambda_i(p_i)$  is suppressed to  $\lambda_i$ .

---

<sup>30</sup>While we do not model the locational choice of drivers within the region, it is fairly easy to see that equidistant positioning location from neighbors is an equilibrium. While there might be other locational choices that might also be equilibria, we focus on the equidistant positioning equilibrium.

Figure 8: Illustration of Circular Model of each Region  $i$  and Driver Allocation



## D Calibration and Identification of the Empirical Model with Demand-Side EOD

This appendix outlines the calibration and identification results of the empirical model with demand-side EOD. The overall strategy is analogous to the model with supply-side EOD. As such, we skip some of the details.

### D.1 Calibration

We adopt a similar approach to the calibration of the supply-side model. Specifically, we estimate the model parameters following the two steps below:

**Step 1** Following eq. (12), We express  $\bar{\lambda}_{id}$  in terms of data and parameters  $(\alpha, \beta)$  as follows:

$$\bar{\lambda}_{id}(\alpha, \beta; \mathcal{D}) = \frac{n_{id}}{(1 - \alpha p_{id})(\frac{c_{id}}{c})(1 - \frac{\beta s_i}{n_{id}})} \quad (16)$$

**Step 2** Given the observables and following step 1, we express the relative outflow between each pair of regions  $i, j$  on each day  $d$  as a function of parameters  $(\alpha, \beta)$ :

$$RO_{ijd} = \frac{A_{id}}{A_{jd}} = \frac{r_i / \bar{\lambda}_{id}(\alpha, \beta; \mathcal{D})}{r_j / \bar{\lambda}_{jd}(\alpha, \beta; \mathcal{D})},$$

We estimate  $(\alpha, \beta)$  by matching the moments of  $RO_{ij}$  for every pair of regions  $i, j$ , where we take the average across  $d \in D$  between the model prediction and the data. The estimates of  $\bar{\lambda}_{id}$  are obtained from  $\bar{\lambda}_{id}(\alpha, \beta, \mathcal{D})$  accordingly.

### D.2 Identification

The identification follows a similar intuition as the supply-side empirical model. Given eq. (16), the identification problem reduces to identify  $\alpha$  and  $\beta$  from the RO moments. Similar to the model with supply-side EOD, we consider the relative outflow across regions for a single day  $d$ . Dropping

the day subscript  $d$ , the model relative outflow can be written as

$$RO_{ij}^{model} = \frac{A_i}{A_j} = \frac{r_i/\bar{\lambda}_i}{r_j/\bar{\lambda}_j} = \frac{\frac{n_i}{c_i}/\bar{\lambda}_i}{\frac{n_j}{c_j}/\bar{\lambda}_j} = \frac{(1 - \alpha p_i)(1 - \beta \frac{s_i}{n_i})}{(1 - \alpha p_j)(1 - \beta \frac{s_i}{n_i})} \quad (17)$$

Following a similar intuition as the empirical model with supply-side EOD, it can be seen that as long as there is sufficient independence between regional price variation and regional density variation, both  $\alpha$  and  $\beta$  are identified. The formal statement and the proof can be found in the appendix.

Our informal identification argument sheds light on why separately identifying supply-side from demand-side EOD is not possible using relative outflows data: variation in density across regions creates geographical skew in access (and hence in relative outflows) through both of the EOD sources. As a result, one cannot separate out how much each source is contributing to the overall economies of density by just observing the relative flows. This is why in calibration with our data, we need to assume (rather than infer) the relative magnitudes of the two sources. That said, as was discussed in the theory section and as will be further elaborated when we report estimation results and counterfactual simulations, our main insights are robust to the source of EOD.

## E Proofs of Propositions in the Theory Sections

This appendix provides proofs for all the results we state for the theory model. It consists of three part. Appendix E.1 gives proofs for the model with supply-side EOD and without inter-region rides (i.e. Proposition 1 through Proposition 3). Appendix E.2 presents proofs for the model with demand-side EOD (i.e. Proposition 4). Appendix E.3 provides the proof for the setting with inter-region rides (i.e. Proposition 5).

Before we get to the proofs for the specific models, we first prove Lemma 1 that speaks to the equilibrium driver equilibrium in general.

**Proof of Lemma 1.** The proof is straightforward. First, observe that, for each region  $i$ ,  $W_i(n_i; p_i)$  must be non-decreasing at  $n_i^*$ . Otherwise, because  $W_i(n_i; p_i)$  is continuous, there exists a small  $\epsilon > 0$  such that  $\frac{c_i}{W_i(n_i^*; p_i)} > \frac{c_i}{W_i(n_i^* + \epsilon; p_i)}$ , the condition in Definition 1 is violated.

Next, we prove the lemma using contradiction. If  $\frac{c_i}{W_i(n_i^*; p_i)} > \bar{c}$ , then given  $W_i(n_i; p_i)$  is continuous and non-decreasing in  $n_i$ , there exists  $\delta > 0$ , such that  $\frac{c_i}{W_i(n_i^*; p_i)} > \frac{c_i}{W_i(n_i^* + \delta; p_i)} > \bar{c}$ . That is, a mass of  $\delta$  drivers will switch from the outside option to driving in region  $i$ . If  $\frac{c_i}{W_i(n_i^*; p_i)} < \bar{c}$ , then the condition in Definition 1 is also violated, so  $n_i^*$  cannot be an equilibrium.

### E.1 Proofs of the theory model with supply-side EOD

Before starting the proofs, we make an observation. With free entry of drivers and abstraction from inter-region rides, the problem in each region is, in some sense, separated from other regions. As such, from the platform's perspective, maximizing the profit over all regions is the equivalent to maximizing the profit of each region independently. Therefore, when solving for the equilibrium

number of drivers for each region, it suffices to look at the equilibrium under the price and wage that maximize the profit of that specific region.

**Proof of Lemma 2, Proposition 1, Proposition 2, and Proposition 3.**

The overall strategy is to prove the lemma and the propositions in a unified framework. First, we prove a series of lemmas that show that, under the parameters  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , the lemma and the propositions can be reduced to a numerical problem. That is, we first show using Lemma A1 - Lemma A5 that each of the propositions holds for a general  $(\bar{\lambda}, t, \bar{c}, \alpha)$  if and only if it holds for  $(\tilde{\lambda}, 1, 1, 1)$ , where  $\tilde{\lambda} = \frac{\bar{\lambda}}{(\bar{c}\alpha)^2t}$ . Showing that the propositions holds for  $(\tilde{\lambda}, 1, 1, 1)$  is essentially numerical, which we carry out accordingly in Lemma A6.

**Lemma A1.** *Assume  $\gamma \in \mathbb{R}$  is a strictly positive number. Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , vectors  $p^*$ ,  $c^*$ , and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha') = (\gamma^2\bar{\lambda}, t, \gamma\bar{c}, \alpha)$ , vectors  $p'^* = p^*$ ,  $c'^* = c^*$ , and  $n'^* = \gamma n^*$  constitute an equilibrium (note that here we are not suppressing the \* notation for equilibria or the regional indices). Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*'} = A_i^*$ .*

**Proof of Lemma A1** Let us expand the notation on platform profits and equilibrium numbers of drivers across the market in a way that allows for explicitly tracking the dependence of all of these on market primitives. To be more specific, we denote the equilibrium number of drivers in region  $i$  under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$  and platform strategy  $(p, c)$  by:

$$n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha)$$

Similarly, we denote the platform profit by:

$$\pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha)$$

Before proving the lemma, we first derive the equilibrium number of drivers given the primitives and the platform strategy. First, note that in equilibrium, the hourly revenue for drivers has to be equal to the reservation value  $\bar{c}$ . Therefore, we have:

$$\frac{c}{W} = \bar{c}$$

Under the model with supply-side EOD, the wait time  $W$  can be expressed in terms of  $n$

$$W(n) = \frac{n}{\lambda} + \frac{t}{n}$$

Replacing  $W$  and then solve for the equilibrium  $n^*$ , we get two solutions:

$$n^* = \frac{\frac{c}{\bar{c}} \mp \sqrt{(\frac{c}{\bar{c}})^2 - \frac{4t}{\lambda(p)}}}{\frac{2}{\lambda(p)}}$$

Note that only the larger solution (i.e., the one with the + sign) is an equilibrium because at the

smaller solution, further driver entry will lead to lower overall wait time, increasing driver revenue. As such, we have:

$$n^* = \frac{\frac{c}{\bar{c}} + \sqrt{\left(\frac{c}{\bar{c}}\right)^2 - \frac{4t}{\lambda(p)}}}{\frac{2}{\lambda(p)}} \quad (18)$$

We now turn to the proof of the lemma, starting with a claim:

**Claim 1.** *For any given  $(p, c)$  where region  $i$  attracts no drivers under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , we have:*

$$\pi_i(p, c | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha) = \gamma^2 \pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha).$$

*Proof of Claim 1.* To see this, note that under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , from eq. (18), region  $i$  will attract no drivers –hence will generate zero profit– if any only if:

$$\left(\frac{c}{\bar{c}}\right)^2 - \frac{4t}{\bar{\lambda}(1-\alpha p)} < 0$$

The corresponding condition under  $(\gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)$  will be:

$$\left(\frac{c}{\gamma \bar{c}}\right)^2 - \frac{4t}{\gamma^2 \bar{\lambda}(1-\alpha p)} < 0$$

The above two conditions are equivalent. This finishes the proof of the claim.  $\square$

Now let us assume  $p$  and  $c$  are such that region  $i$  does get a positive number of drivers. There, the number of drivers present in  $i$  under  $(\bar{\lambda}, t, \bar{c}, \alpha)$  will be uniquely give by:

$$n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha) = \frac{\frac{c_i}{\bar{c}} + \sqrt{\left(\frac{c_i}{\bar{c}}\right)^2 - \frac{4t_i}{\bar{\lambda}_i(1-\alpha p_i)}}}{\frac{2}{\bar{\lambda}_i(1-\alpha p_i)}}$$

The same quantity under  $(\gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)$  will be uniquely give by:

$$\begin{aligned} n_i^*(p, c | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha) &= \frac{\frac{c_i}{\gamma \bar{c}} + \sqrt{\left(\frac{c_i}{\gamma \bar{c}}\right)^2 - \frac{4t_i}{\gamma^2 \bar{\lambda}_i(1-\alpha p_i)}}}{\frac{2}{\gamma^2 \bar{\lambda}_i(1-\alpha p_i)}} \\ &= \gamma \frac{\frac{c_i}{\bar{c}} + \sqrt{\left(\frac{c_i}{\bar{c}}\right)^2 - \frac{4t_i}{\bar{\lambda}_i(1-\alpha p_i)}}}{\frac{2}{\bar{\lambda}_i(1-\alpha p_i)}} = \gamma n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha) \end{aligned}$$

Also note that the equilibrium driver wait time  $W_i$  in region  $i$  under  $(\bar{\lambda}, t, \bar{c}, \alpha)$  has to be  $\frac{c_i}{\bar{c}}$  in order to equate driver revenues in the region with the reservation value. Similarly, under  $(\gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)$ , the driver wait time in region  $i$  will be  $\frac{c_i}{\gamma \bar{c}}$ .

Next, note that for any  $p, c$ :

$$\pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha) = (p_i - c_i) \times \frac{n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha)}{\frac{c_i}{\bar{c}}}$$

Also:

$$\begin{aligned}\pi_i(p, c | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha) &= (p_i - c_i) \times \frac{n_i^*(p, c | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)}{\frac{c_i}{\gamma \bar{c}}} \\ &= (p_i - c_i) \times \frac{\gamma n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha)}{\frac{c_i}{\gamma \bar{c}}} = \gamma^2 \pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha)\end{aligned}$$

Therefore, the profit functions under  $(\bar{\lambda}, t, \bar{c}, \alpha)$  and  $(\gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)$  are fully proportional in each region, which mean they will have the exact same maximizers  $p^*$  and  $c^*$ . Given these maximizers, we will have the number of drivers in each region  $i$  under  $(\gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha)$  will be given by:

$$n_i^{*'} \equiv n_i^*(p^*, c^* | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha) = \gamma n_i^*(p^*, c^* | \bar{\lambda}, t, \bar{c}, \alpha) \equiv \gamma n_i^*$$

Finally:

$$A_i^{*'} = \frac{n_i^{*'}}{\left(\frac{c_i^*}{\gamma \bar{c}}\right) \times \gamma^2 \bar{\lambda}_i} = \frac{\gamma n_i^*}{\left(\frac{c_i^*}{\gamma \bar{c}}\right) \times \gamma^2 \bar{\lambda}_i} = \frac{n_i^*}{\left(\frac{c_i^*}{\bar{c}}\right) \times \bar{\lambda}_i} = A_i^*$$

The proof of the lemma is now complete.  $\square$

**Lemma A2.** *Assume  $\gamma \in \mathbb{R}$  is a strictly positive number. Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , vectors  $p^*$ ,  $w^*$ , and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha') = (\bar{\lambda}, t, \gamma \bar{c}, \frac{\alpha}{\gamma})$ , vectors  $p^{*'} = \gamma p^*$ ,  $c^{*'} = \gamma c^*$ , and  $n^{*'} = n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*'} = A_i^*$ .*

**Proof of Lemma A2.** The steps of the proof for this lemma are very similar to those in the proof of Lemma A1. Hence, we skip the details.  $\square$

We just point out that there is a simple intuition for this result: a change of primitives from  $(\bar{\lambda}, t, \bar{c}, \alpha)$  to  $(\bar{\lambda}, t, \gamma \bar{c}, \frac{\alpha}{\gamma})$  should not be expected to change the equilibria because it is effectively a “currency change.” The only two places where “money” becomes involved in the model are where demand is determined based on the price of a ride and where supply is determined based on wage. If the reservation hourly revenue is multiplied by  $\gamma$  and, at the same time, the price sensitivity of passengers is divided by the same factor  $\gamma$ , it is as if we are looking at the same market but with a new currency one unit of which is worth  $\frac{1}{\gamma}$  times one unit of the old currency. This should not change the equilibrium supply, demand, and access. It should only multiply the equilibrium prices and wages by  $\gamma$ .

**Lemma A3.** *Assume  $\gamma \in \mathbb{R}$  is a strictly positive number. Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , vectors  $p^*$ ,  $w^*$ , and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha') = (\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha)$ , vectors  $p^{*'} = p^*$ ,  $c^{*'} = c^*$ , and  $n^{*'} = \gamma n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*'} = A_i^*$ .*

**Proof of Lemma A3.** Again, the steps of this proof are very similar to those in the previous two lemmas. As such, we skip the details.  $\square$

Our next lemma combines the previous three.

**Lemma A4.** Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , vectors  $p^*$ ,  $c^*$ , and  $n^*$  constitute an equilibrium if and only if under primitives  $(\tilde{\lambda}, 1, 1, 1)$  where  $\tilde{\lambda}_i = \frac{\bar{\lambda}_i}{(\bar{c}\alpha)^2 t_i}$  for each  $i$ , the following vectors constitute an equilibrium:  $p^{*'} = \alpha p^*$ ,  $c^{*'} = \alpha c^*$ , and  $n^{*'} = \frac{1}{(\bar{c}\alpha)t} n^*$ . Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*'} = A_i^*$ .

**Proof of Lemma A4.** Apply Lemma A1 with  $\gamma = \frac{1}{\bar{c}\alpha}$  on primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$ . On the resulting primitives, apply Lemma A2 with  $\gamma = \alpha$ . Finally, on the resulting primitives, apply Lemma A3 with  $\gamma = \frac{1}{t}$ . This will establish the equivalence claimed in this lemma.  $\square$

With this result in hand, we now turn to the main two lemmas that prove Lemma 2 and Proposition 1 through Proposition 3.

**Lemma A5.** Lemma 2 and Proposition 1 through Proposition 3 hold under all generic primitives  $(\bar{\lambda}, t, \bar{c}, \alpha)$  if they hold under primitives in the form of  $(\tilde{\lambda}, 1, 1, 1)$ .

**Proof of Lemma A5.** Suppose Lemma 2 and Proposition 1 through Proposition 3 hold under primitives that take the form of  $(\tilde{\lambda}, 1, 1, 1)$ . We show it holds for a general  $(\bar{\lambda}, t, \bar{c}, \alpha)$ . To this end, choose vector  $\tilde{\lambda}$  such that  $\forall i : \tilde{\lambda}_i = \frac{\bar{\lambda}_i}{(\bar{c}\alpha)^2 t_i}$ . Next, we prove each lemma/proposition for general primitives.

*Lemma 2:*

Given that Lemma 2 holds under  $(\tilde{\lambda}, 1, 1, 1)$ , there exists a unique equilibrium under  $(\tilde{\lambda}, 1, 1, 1)$ . Denote it  $(\tilde{p}, \tilde{c}, \tilde{n})$ . By Lemma A4, there also exists equilibrium  $(p^*, c^*, n^*)$  under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , which is given by  $n^* = \tilde{n}_i (\bar{c}\alpha t_i)$  for each  $i$ . And  $\tilde{n}$  being unique implies that  $n^*$  is unique. Also, for any  $i$  for which  $n^* > 0$ , we know  $\tilde{n} > 0$ . From the lemma, we know there are unique equilibrium price  $\tilde{p}_i$  and wage  $\tilde{c}_i$ . Then  $p^*$  and  $c^*$  are also unique.

*Proposition 1:*

First we show the first half of the proposition that speaks to  $n_i^* \neq 0 \neq n_j^*$  holds for the generic primitives. By assumption, this statement holds under  $(\tilde{\lambda}, 1, 1, 1)$ . By the construction of  $\tilde{\lambda}$ , it is immediate that:

$$\tilde{\lambda}_i > \tilde{\lambda}_j \Leftrightarrow \frac{\bar{\lambda}_i}{t_i} > \frac{\bar{\lambda}_j}{t_j}$$

Also by Lemma A4, we know that equilibrium access to rides in each region  $i$  is the same under  $(\tilde{\lambda}, 1, 1, 1)$  and  $(\bar{\lambda}, t, \bar{c}, \alpha)$ . It immediately follows that the first half of the statement holds under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ .

For the second half of the proposition that speaks to the regions without supply, by assumption, we know there is a  $\mu$  such that:

$$\forall i : \frac{\tilde{\lambda}_i}{1} < \mu \Leftrightarrow \tilde{n}_i = 0$$

Note that by construction,  $\tilde{\lambda}_i < \mu$  is equivalent to  $\frac{\bar{\lambda}_i}{t_i} < \mu(\bar{c}\alpha)^2$ . Also note that  $n_i^* = 0$  if and only if  $\tilde{n}_i = 0$ . Therefore, we have:

$$\forall i : \frac{\bar{\lambda}_i}{t_i} < \mu(\bar{c}\alpha) \Leftrightarrow n_i^* = 0$$

This finishes the proof of Proposition 1 under the generic primitives.

*Proposition 2:*

Key to proving the four statements of Proposition 2 is the observation that the same equivalence established between  $(\tilde{\lambda}, 1, 1, 1)$  and  $(\bar{\lambda}, t, \bar{c}, \alpha)$  by Lemma A4 may also be established between  $(\gamma\tilde{\lambda}, 1, 1, 1)$  and  $(\gamma\bar{\lambda}, t, \bar{c}, \alpha)$  for any positive real  $\gamma$ . Even the multipliers  $\frac{1}{(\bar{c}\alpha)^2 t}$  remain the same. With this observation, we turn to the proofs.

*Statement 1.* By Lemma A4, We know  $\tilde{p} = \alpha p^*$ ,  $\tilde{c} = \alpha c$ , and  $\tilde{n} = \frac{1}{(\bar{c}\alpha)^2 t} n^*$ . By the first statement being true for  $(\bar{\lambda}, 1, 1, 1)$ , there exists equilibrium  $(\tilde{c}', \tilde{p}', \tilde{n}')$  under the primitives  $(\gamma\tilde{\lambda}, 1, 1, 1)$ , where  $\tilde{n}'$  is unique. Also  $(\tilde{p}', \tilde{c}')$  are unique for  $\tilde{n}' > 0$ . Then by Lemma A4, there exists equilibrium  $(p^{*'}, c^{*'}, n^{*'})$  under the primitives  $(\gamma\bar{\lambda}, t, \alpha, \bar{c})$ , with  $n^{*'} = (\bar{c})\alpha t \tilde{n}'$ ,  $p^{*'} = \frac{1}{\alpha}\tilde{p}$ , and  $c^{*'} = \frac{1}{\alpha}\tilde{c}$ . Therefore  $n^{*'}$  is unique, and  $(p^{*'}, c^{*'})$  unique for  $n^{*'} > 0$ .

*Statement 2.* By Lemma A4, we know  $n_i^* > 0 \Rightarrow \tilde{n}_i > 0$ . By the proposition being true for  $(\bar{\lambda}, 1, 1, 1)$ , we know that  $\tilde{n}_i > 0 \Rightarrow \tilde{n}'_i > 0$  where  $\tilde{n}'$  is the equilibrium driver allocation under  $(\gamma\tilde{\lambda}, 1, 1, 1)$ . Finally, again by Lemma A4, we have  $\tilde{n}'_i > 0 \Rightarrow n_i^{*'} > 0$ . This finishes the proof of this statement.

*Statement 3.* Lemma A4 shows that access to rides in all regions is preserved under the transformation from  $(\bar{\lambda}, t, \bar{c}, \alpha)$  to  $(\tilde{\lambda}, 1, 1, 1)$  and the transformation from  $(\gamma\bar{\lambda}, t, \bar{c}, \alpha)$  to  $(\gamma\tilde{\lambda}, 1, 1, 1)$ . This, together with  $\tilde{\lambda}_i \geq \tilde{\lambda}_j \Leftrightarrow \frac{\tilde{\lambda}_i}{t_i} \geq \frac{\tilde{\lambda}_j}{t_j}$  and the assumption that the proposition holds under  $(\tilde{\lambda}, 1, 1, 1)$  gives the proof of this statement under  $(\gamma\bar{\lambda}, t, \bar{c}, \alpha)$ .

*Statement 4.* The logic for the proof of this statement is identical to that for the proof of the previous one.

*Proposition 3:*

Note that by construction of  $\tilde{\lambda}$  and by Lemma A4, for any  $i, j$  the following hold:

1.  $\tilde{\lambda}_i \geq \tilde{\lambda}_j \Leftrightarrow \frac{\tilde{\lambda}_i}{t_i} \geq \frac{\tilde{\lambda}_j}{t_j}$  and a similar statement when both inequalities are strict.
2.  $\tilde{c}_i \leq \tilde{c}_j \Leftrightarrow c_i^* \leq c_j^*$  and a similar statement when both inequalities are strict.
3.  $\tilde{p}_i \leq \tilde{p}_j \Leftrightarrow p_i^* \leq p_j^*$  and a similar statement when both inequalities are strict.
4.  $\tilde{p}_i - \tilde{c}_i \leq \tilde{p}_j - \tilde{c}_j \Leftrightarrow p_i^* - c_i^* \leq p_j^* - c_j^*$  and a similar statement when both inequalities are strict.

These four statements, together with the assumption that Proposition 3 of the proposition on economies of density holds under  $(\tilde{\lambda}, 1, 1, 1)$  imply that it also holds under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ .

The proof of Lemma A5 is now complete.  $\square$

Our last step in proving Lemma 2 and Proposition 1 through Proposition 3 is to show that it does indeed hold under  $(\tilde{\lambda}, 1, 1, 1)$ . The convenient feature of this remaining step is that it can be done fully numerically. The following lemma takes on this task.

**Lemma A6.** *Lemma 2 and Proposition 1 through Proposition 3 holds if  $(\bar{\lambda}, t, \bar{c}, \alpha)$  takes the specific form of  $(\tilde{\lambda}, 1, 1, 1)$ .*

**Proof of Lemma A6.** We start by proving the second part of Proposition 1 that gives conditions for a region not to be supplied, and then move to all others.

*Second half of Proposition 1.* Here, we are looking for conditions on potential demand  $\bar{\lambda}_i$  such that the platform can get a positive number of drivers into region  $i$  without sustaining any loss. Note that from our analysis before, we know the region can get a positive number of drivers if:

$$\left(\frac{c_i}{\bar{c}}\right)^2 \geq \frac{4t_i}{\bar{\lambda}_i(1-\alpha p_i)}$$

From  $\bar{c} = t_i = \alpha = 1$  and  $\bar{\lambda}_i = \tilde{\lambda}_i$ , this turns into:

$$c_i^2 \geq \frac{4}{\tilde{\lambda}_i(1-p_i)}$$

Or, equivalently:

$$\tilde{\lambda}_i \geq \frac{4}{c_i^2(1-p_i)}$$

As such, the minimum requirement volume for  $\tilde{\lambda}_i$  depends on the maximum possible value for  $c_i^2(1-p_i)$  subject to the constraint that the platform does not run any losses. This happens when  $p_i = c_i$ . Therefore, the region will get a positive number of drivers if and only if:

$$\tilde{\lambda}_i \geq \frac{4}{\max_{c_i} c_i^2(1-c_i)} = 27$$

This says each region  $i$  gets a positive number of drivers if and only if  $\tilde{\lambda}_i \geq 27$  which proves the second half of Proposition 1.

*Existence of equilibrium in Lemma 2.* Now we prove the statement on the existence of the equilibrium in Lemma 2. The proof above shows that if  $\tilde{\lambda} < 27$ , then the maximum profit is always 0 and it is trivial that  $(p^*, c^*)$  exists and  $n^* = 0$  exists for  $\tilde{\lambda} < 27$ . As such, we focus on the case where  $\tilde{\lambda} \geq 27$ . We show the existence by dividing the parameter space  $\mathbf{R}^2$  for  $(p, c)$  into four subsets and show that the maximum of the profit function exists in each case. Then combining all cases, there exists  $(p^*, c^*)$  and corresponding  $n^*$  that maximizes the profit function.

Case (1): If  $c_i \leq 0$ , then drivers receive zero revenue and no drivers operate in the platform and no rides realized. Thus the equilibrium is  $p_i^* \in \mathbf{R}$ ,  $c_i^* \leq 0$ ,  $n_i^* = 0$ , which gives  $\pi_i^* = 0$ .

Case (2): If  $p_i \geq 1$ , then there is no demand and no rides are realized. Drivers receive zero revenue from the platform and thus the equilibrium is achieved for  $p_i^* \geq 1$ ,  $c_i^* \in \mathbf{R}$ ,  $n_i^* = 0$ , which gives  $\pi_i^* = 0$ .

Now we discuss the remaining parameter space for  $(p, c)$  where  $p_i < 1$  and  $c_i > 0$ . Recall that the wait time as a function of number of drivers is given by

$$W(n_i) = \frac{n_i}{\tilde{\lambda}_i(1-p_i)} + \frac{1}{n_i}$$

The function  $W(n_i)$  (with  $n_i > 0$ ) is decreasing in  $(0, \sqrt{\tilde{\lambda}_i(1-p_i)})$ , increasing in  $(\sqrt{\tilde{\lambda}_i(1-p_i)}, \infty)$  and minimized at  $n_i = \sqrt{\tilde{\lambda}_i(1-p_i)}$ . The minimum is  $W(\sqrt{\tilde{\lambda}_i(1-p_i)}) = \frac{2}{\sqrt{\tilde{\lambda}_i(1-p_i)}}$ .

Case (3): If  $p_i < 1$  and  $0 < c < \frac{2}{\sqrt{\tilde{\lambda}_i(1-p_i)}}$ , then  $c \leq \min_{n_i > 0} W(n_i)$ . Therefore,  $\frac{c}{W(n_i)} < 1 = \bar{c}, \forall n_i$ . That is, the driver revenue is always lower than the reservation wage. So  $n_i^* = 0$  and  $\pi_i^* = 0$ . The equilibrium is achieved with  $(p_i^*, c_i^*) \in \{(p, c) : p < 1, 0 < c < \frac{2}{\sqrt{\tilde{\lambda}_i(1-p_i)}}\}$  and  $n_i^* = 0$ , which gives  $\pi_i^* = 0$ .

Case (4): If  $p < 1$  and  $c \geq \frac{2}{\sqrt{\tilde{\lambda}_i(1-p_i)}}$ , then as the roof of Lemma A1 shows, only the larger solution of the equation  $W(n_i) = c$  is the equilibrium. The equilibrium number of drivers is given by

$$n_i^*(p_i, c_i) = \frac{c_i + \sqrt{c_i^2 - \frac{4}{\tilde{\lambda}(1-p_i)}}}{\frac{2}{\tilde{\lambda}(1-p_i)}}$$

And the profit function is given by

$$\pi_i(p_i, c_i) = (p_i - c_i) \frac{n_i(p_i, c_i)}{c_i}$$

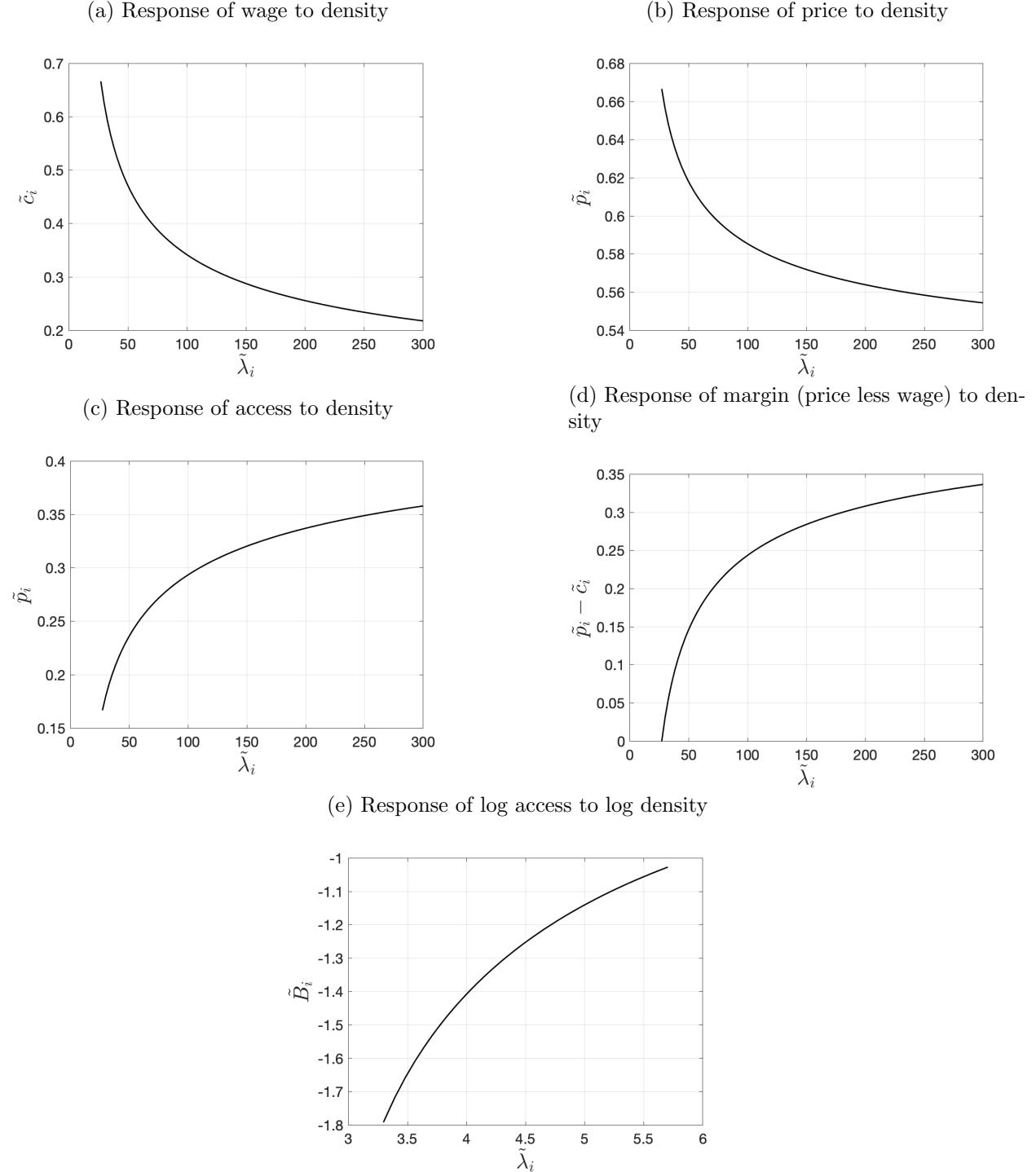
Considering that in all previous cases we have discussed above, it is at least feasible to achieve zero profit. Therefore, to find the maximum, we focus on the parameter space that leads to non-negative profit. As such, we focus on the  $(p_i, c_i)$  such that  $p \geq c$  and consider the set  $\mathcal{X} = \{(p_i, c_i) : 0 \leq p_i \leq 1, c_i \leq p_i, c_i \geq \frac{2}{\sqrt{\tilde{\lambda}_i(1-p_i)}}\}$ . The set is non-empty given  $\tilde{\lambda} \geq 27$ .

Given that the profit function is continuous and the set  $\mathcal{X}$  is closed and bounded, there exists  $(p_i^*, c_i^*)$  that maximizes the profit function. And it can be seen that the maximum is non-negative (e.g. pick any  $(p_i, c_i) \in \mathbf{X}$  with  $p_i = c_i$ ). Therefore, combining all the cases above, the maximum of the profit function always exists, and thus  $(p^*, c^*, n^*)$  exist.

Next, we prove the remaining part of Lemma 2 and Proposition 1 through Proposition 3 numerically. To do this, we need to study the equilibrium price  $\tilde{p}_i$ , equilibrium wage  $\tilde{c}_i$ , equilibrium number of drivers  $\tilde{n}_i$ , and equilibrium access  $\tilde{A}_i$  in a given region  $i$  as a function of  $\tilde{\lambda}_i$ . The key is that all of the other parameters are equal to 1 and  $\tilde{\lambda}_i$  is the only parameter changing. Therefore, for any value that  $\tilde{\lambda}_i$  assumes, we have a fully numerical problem that can be solved using a software such as R. Below, we are providing graphs of these equilibrium quantities as functions of  $\tilde{\lambda}_i$ . Fig. 9 provides these results. First, the figure gives the unique optimal price and wage computed for  $\tilde{\lambda}_i \geq 27$ , which shows that Lemma 2 holds. Second, access  $\tilde{A}_i$  is strictly increasing in density. This means if two regions  $i, j$  are such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a lower access to rides in the equilibrium. This completes the proof of Proposition 1. Additionally, as shown in the figure, optimal price  $\tilde{p}_i$  and optimal wage  $\tilde{c}_i$  both strictly decrease in density  $\tilde{\lambda}_i$ . This means if two regions  $i, j$  are such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a higher price and a higher wage in the equilibrium. Also margin  $\tilde{p}_i - \tilde{c}_i$  is strictly increasing in density. This means if two regions  $i, j$  are

such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a lower margin in the equilibrium. These shows that Proposition 3 holds.

Figure 9: Numerical results that are necessary to prove Lemma A6



As for Proposition 2, statement 1 follows directly from the second half of Proposition 1. To

prove statement 2, first note that the second part of this statement (i.e.,  $\tilde{\lambda}_i > \tilde{\lambda}_j \Rightarrow \frac{A_j(n_j^{*'})}{A_i(n_i^{*'})} < 1$ ) is implied by the first half of Proposition 1, which we have already proved. We, hence, will only prove:

$$\tilde{\lambda}_i > \tilde{\lambda}_j \Rightarrow \frac{A_j(n_j^*)}{A_i(n_i^*)} < \frac{A_j(n_j^{*'})}{A_i(n_i^{*'})}$$

In order to prove this statement, we first state and prove a lemma. Recall, again, that given the optimization problems in different regions are separate from each other, the equilibrium access  $A_i(n_i^*)$  is only a function of  $\tilde{\lambda}_i, p_i, c_i$  rather than those at other regions. Therefore, the notation  $i$  may be suppressed. In the proofs of some of the last statements we have even suppressed the notation on  $n^*$  and denote the equilibrium access in region  $i$  simply by  $A$ . In our next lemma, we will keep suppressing those notations. We also note that access has been an implicit function of model parameters, in particular  $\tilde{\lambda}_i$ . In the proof of this lemma, we make the exposition of the dependence of equilibrium access  $A$  on parameter  $\tilde{\lambda}_i$  explicit by denoting the equilibrium access as  $A(\tilde{\lambda}_i)$ . In other words,  $A(\tilde{\lambda}_i)$  is the equilibrium access in a region with demand  $\tilde{\lambda}_i$  when the platform has set the optimal wage  $c^*$  and price  $p^*$  given density  $\tilde{\lambda}_i$  and drivers respond accordingly.

**Lemma A7.** Define function  $B(\tau)$  as

$$B(\tau) \equiv \log(A(e^\tau))$$

The second statement of Proposition 2 holds if  $B$  is a strictly concave function whenever defined.

**Proof of Lemma A7.** First notice that  $B$  is defined whenever  $A > 0$  which is whenever  $\tau$  is larger than some cutoff.

Next, take two regions  $i, j$  with  $\tilde{\lambda}_i > \tilde{\lambda}_j$ . Set  $\tau_i = \log(\tilde{\lambda}_i)$  and  $\tau_j = \log(\tilde{\lambda}_j)$ . Also set  $\beta = \log(\gamma)$  where  $\gamma$  is the scalar larger than one in the statement of the proposition. By strict concavity of  $B$ , we have:

$$B(\tau_i + \beta) - B(\tau_i) < B(\tau_j + \beta) - B(\tau_j)$$

This implies:

$$\begin{aligned} \log[A(e^{\tau_i+\beta})] - \log[A(e^{\tau_i})] &< \log[A(e^{\tau_j+\beta})] - \log[A(e^{\tau_j})] \\ \Rightarrow \log\left[\frac{A(e^{\tau_i+\beta})}{A(e^{\tau_i})}\right] &< \log\left[\frac{A(e^{\tau_j+\beta})}{A(e^{\tau_j})}\right] \\ \Rightarrow \frac{A(\gamma\tilde{\lambda}_i)}{A(\tilde{\lambda}_i)} &< \frac{A(\gamma\tilde{\lambda}_j)}{A(\tilde{\lambda}_j)} \\ \Rightarrow \frac{A(\tilde{\lambda}_j)}{A(\tilde{\lambda}_i)} &< \frac{A(\gamma\tilde{\lambda}_j)}{A(\gamma\tilde{\lambda}_i)} \end{aligned}$$

which is exactly the statement we needed to prove. This completes the proof of the lemma.  $\square$

Therefore, to prove the second statement of Proposition 2 the only thing we need to prove is strict concavity of the function  $B$ . This can be seen from the last panel in Fig. 9 which shows that log of access is concave in log of density.

Finally, statement 3 of Proposition 2 is straightforward and left to the reader (the key would be to note that as the market gets sufficiently large, pickup times in all regions become negligible; hence the pricing problem in all regions boils down to that in a model without economies of density, giving all regions the same price, wage, and access to rides).

This finishes the proof of Lemma A6 and, hence, the proofs for all lemma/propositions of the theory model with supply-side EOD (i.e. Lemma 2 and Proposition 1 through Proposition 3). ■

## E.2 Proofs of the theory model with demand-side EOD

This appendix proves Proposition 4, which states that an equivalence of Lemma 2 and Proposition 1 through Proposition 3 holds under the theory model with demand-side EOD. The proof strategy is largely similar to that for the model with supply-side EOD, as discussed in Appendix E.1. We prove the propositions in a unified framework, where the propositions can be reduced to a numerical problem. Specifically, Lemma A8 presents similar statements to Lemma A1 through Lemma A3 and shows the sets of primitives that leads to equivalent equilibrium. Lemma A9 and Lemma A10 resemble Lemma A4 and Lemma A5 and states that the proposition can be reduced to a numerical problem, where the proposition holds if it holds for a specific set of primitives. Finally, Lemma A11 shows that the proposition does hold for the set of primitives.

**Lemma A8.** *Assume  $\gamma \in \mathbb{R}$  is a strictly positive number. Then the following statements hold: is equivalent to any of the following statements:*

1. *Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , vectors  $p^*$ ,  $c^*$  and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha', \beta') = (\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)$ , vectors  $p^{*\prime} = p^*$ ,  $c^{*\prime} = c^*$  and  $n^{*\prime} = \gamma n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*\prime} = A_i^*$ .*
2. *Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , vectors  $p^*$ ,  $c^*$  and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha', \beta') = (\frac{1}{\gamma} \bar{\lambda}, t, \bar{c}, \frac{1}{\gamma} \alpha, \beta)$ , vectors  $p^{*\prime} = \gamma p^*$ ,  $c^{*\prime} = \gamma c^*$  and  $n^{*\prime} = n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*\prime} = A_i^*$ .*
3. *Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , vectors  $p^*$ ,  $c^*$  and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}, \frac{1}{\gamma} t, \bar{c}, \alpha, \gamma \beta)$ , vectors  $p^{*\prime} = p^*$ ,  $c^{*\prime} = c^*$  and  $n^{*\prime} = n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*\prime} = A_i^*$ .*
4. *Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , vectors  $p^*$ ,  $c^*$  and  $n^*$  constitute an equilibrium if and only if under primitives  $(\gamma \bar{\lambda}, t, \gamma \bar{c}, \alpha, \beta)$ , vectors  $p^{*\prime} = p^*$ ,  $c^{*\prime} = c^*$  and  $n^{*\prime} = n^*$  constitute an*

equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  
 $\forall i : A_i^{*'} = A_i^*$ .

*Proof.* As the proofs of the statements are very similar, we prove the first statement and skip the rest.

Under the model with demand-side EOD, we denote the equilibrium number of drivers in region  $i$  under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  and platform strategy  $(p, c)$  by:

$$n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta)$$

Similarly, we denote the platform profit of region  $i$  by:

$$\pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta)$$

Note that the equilibrium number of drivers in each region is given by

$$\frac{n}{\bar{\lambda}(1-\alpha p)(1-\beta \frac{t}{n})} = \frac{c}{\bar{c}}$$

Solving the equation, we get two solutions:

$$n^* = \frac{1 \mp \sqrt{1 - \frac{4\beta t}{\frac{c}{\bar{c}}\bar{\lambda}(1-\alpha p)}}}{\frac{2}{\frac{c}{\bar{c}}\bar{\lambda}(1-\alpha p)}}$$

The equilibrium number of drivers  $n^*$  takes real value if and only if  $4\beta t \leq \frac{c}{\bar{c}}\bar{\lambda}(1-\alpha p)$ . Now we show that the region cannot get a positive number of drivers if  $4\beta t_i > \frac{c_i}{\bar{c}}\bar{\lambda}_i(1-\alpha p_i)$ .

The hourly wage for the rider is

$$\frac{c_i}{\frac{n_i}{\bar{\lambda}_i(1-\alpha p_i)(1-\frac{\beta t_i}{n_i})}} = \frac{c_i \bar{\lambda}_i(1-\alpha p_i)(1-\frac{\beta t_i}{n_i})}{n_i} < \frac{4\beta t_i(1-\frac{\beta t_i}{n_i})}{n_i} \bar{c} = \frac{4\beta t_i n_i - 4\beta^2 t_i^2}{n_i^2} \bar{c}$$

Given  $(n_i - 2\beta t_i)^2 \geq 0$ , we have  $\frac{4\beta t_i n_i - 4\beta^2 t_i^2}{n_i^2} < 1$ . So when  $\frac{c_i}{\bar{c}}\bar{\lambda}_i(1-\alpha p_i) < 4\beta t_i$ , the hourly wage for working on the ride sharing platform is lower than  $\bar{c}$ . Therefore, the region cannot get a positive number of drivers.

Note that only the larger solution (i.e., the one with the + sign) is an equilibrium because at the lower solution, further driver entry will lead to lower overall wait time, increasing driver revenue. As such, we have:

$$n^* = \frac{1 + \sqrt{1 - \frac{4\beta t}{\frac{c}{\bar{c}}\bar{\lambda}(1-\alpha p)}}}{\frac{2}{\frac{c}{\bar{c}}\bar{\lambda}(1-\alpha p)}} \quad (19)$$

We now turn to the proof of the lemma, starting with a claim:

**Claim 2.** For any given  $(p, c)$  where region  $i$  attracts no drivers under  $(\bar{\lambda}, t, \bar{c}, \alpha)$ , we have:

$$\pi_i(p, c | \gamma^2 \bar{\lambda}, t, \gamma \bar{c}, \alpha) = \gamma \pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha).$$

*Proof of Claim 2.* To see this, note that under  $(\gamma \bar{\lambda}, \gamma t, \alpha, \beta, \bar{c})$ , from eq. (19), region  $i$  will attract no drivers –hence will generate zero profit– if any only if:

$$4\beta t > \frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)$$

The corresponding condition under  $(\gamma \bar{\lambda}, \gamma t, \alpha, \beta, \bar{c})$  will be:

$$4\beta \gamma t > \frac{c}{\bar{c}} \gamma \bar{\lambda} (1 - \alpha p)$$

The above two conditions are equivalent. This finishes the proof of the claim.  $\square$

Now let us assume  $p$  and  $c$  are such that region  $i$  does get a positive number of drivers. There, the number of drivers present in  $i$  under  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  will be uniquely give by:

$$n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta) = \frac{1 + \sqrt{1 - \frac{4\beta t}{\frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)}}}{\frac{2}{\frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)}}$$

The same quantity under  $(\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)$  will be uniquely give by:

$$n_i^*(p, c | \gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta) = \frac{1 + \sqrt{1 - \frac{4\beta t}{\frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)}}}{\frac{2}{\frac{c}{\bar{c}} \gamma \bar{\lambda} (1 - \alpha p)}} = \gamma \frac{1 + \sqrt{1 - \frac{4\beta t}{\frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)}}}{\frac{2}{\frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p)}} = \gamma n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta)$$

Also note that the equilibrium driver wait time  $W_i$  in region  $i$  under  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  has to be  $\frac{c_i}{\bar{c}}$  in order to equate driver revenues in the region with the reservation value. Similarly, under  $(\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)$ , the driver wait time in region  $i$  will be  $\frac{c_i}{\bar{c}}$ .

Next, note that for any  $p, c$ :

$$\pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha) = (p_i - c_i) \times \frac{n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta)}{\frac{c_i}{\bar{c}}}$$

Also:

$$\begin{aligned} \pi_i(p, c | \gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta) &= (p_i - c_i) \times \frac{n_i^*(p, c | \gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)}{\frac{c_i}{\bar{c}}} \\ &= (p_i - c_i) \times \frac{\gamma n_i^*(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta)}{\frac{c_i}{\bar{c}}} = \gamma \pi_i(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta) \end{aligned}$$

Therefore, the profit functions under  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  and  $(\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)$  are fully proportional in each region, which mean they will have the exact same maximizers  $p^*$  and  $c^*$ . Given these maximizers, we will have the number of drivers in each region  $i$  under  $(\gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta)$  will be given

by:

$$n_i^{*'} \equiv n_i^*(p^*, c^* | \gamma \bar{\lambda}, \gamma t, \bar{c}, \alpha, \beta) = \gamma n_i^*(p^*, c^* | \bar{\lambda}, t, \bar{c}, \alpha, \beta) \equiv \gamma n_i^*$$

Finally:

$$A_i^{*'} = \frac{n_i^{*'}}{\left(\frac{c_i^*}{\bar{c}}\right) \times \gamma \bar{\lambda}_i} = \frac{\gamma n_i^*}{\left(\frac{c_i^*}{\bar{c}}\right) \times \gamma \bar{\lambda}_i} = \frac{n_i^*}{\left(\frac{c_i^*}{\bar{c}}\right) \times \bar{\lambda}_i} = A_i^*$$

The proof of the first statement of the lemma is now complete.  $\square$

The next lemma combines the statements of the previous one.

**Lemma A9.** *Under primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , vectors  $p^*$ ,  $c^*$ , and  $n^*$  constitute an equilibrium if and only if under primitives  $(\bar{\lambda}', t', \bar{c}', \alpha', \beta') = (\frac{1}{\alpha \beta \bar{c} t} \bar{\lambda}, 1, 1, 1, 1)$ , vectors  $p^{*'} = \alpha p^*$ ,  $c^{*'} = \alpha c^*$ ,  $n^{*'} = \frac{1}{\beta t} n^*$  constitute an equilibrium. Additionally, access to rides in all regions are equal under these two equilibria:  $\forall i : A_i^{*'} = A_i^*$ .*

*Proof.* Apply part 3 of Lemma Lemma A8 with  $\gamma = \frac{1}{\beta}$ , then part 1 of Lemma Lemma A8 with  $\gamma = \frac{1}{\beta t}$ , part 2 of Lemma Lemma A8 with  $\gamma = \alpha$ , and then part 4 of Lemma Lemma A8 with  $\gamma = \frac{1}{\bar{c}}$ .

With the previous lemma in hand, as in the proofs for the model with supply-side EOD, we turn to the two main lemmas that prove Proposition 4.

**Lemma A10.** *Proposition 4 holds under all generic primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  if it holds under primitives in the form of  $(\tilde{\lambda}, 1, 1, 1, 1)$ .*

*Proof of Lemma A10.* The proof is very similar to Lemma A5 with the model with supply-side EOD. Suppose Proposition 4 holds under primitives in the form of  $(\tilde{\lambda}, 1, 1, 1, 1)$ . That is, Lemma 2 and Proposition 1 through Proposition 3 hold under the model with demand-side EOD. We show that, still with the demand-side EOD model, the lemma and the propositions hold for a general  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ . To this end, choose vector  $\tilde{\lambda}$  such that  $\tilde{\lambda} = \frac{1}{\alpha \beta \bar{c} t} \bar{\lambda}$ . Next, we prove each lemma/proposition for general primitives. We suppress the regional indices below.

*Lemma 2:*

Given that Lemma 2 holds under  $(\tilde{\lambda}, 1, 1, 1, 1)$ , there exists equilibrium  $(\tilde{p}, \tilde{c}, \tilde{n})$  under  $(\tilde{\lambda}, 1, 1, 1, 1)$ . By Lemma A9, there also exists equilibrium  $(p^*, c^*, n^*)$  under  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , which is given by  $p^* = \frac{1}{\alpha} \tilde{p}$ ,  $c^* = \frac{1}{\alpha} \tilde{c}$  and  $n^* = (\beta t) \tilde{n}$ . Then for  $n^* > 0$ , we have  $\tilde{n}_i > 0$ . Then  $p^*$  and  $c^*$  are unique if  $n^* > 0$  because  $\tilde{n}_i > 0$  and  $(\tilde{p}, \tilde{c})$  are unique.

*Proposition 1:*

First we show the first half of the proposition that speaks to  $n_i^* \neq 0 \neq n_j^*$  holds for the generic primitives. By assumption, this statement holds under  $(\tilde{\lambda}, 1, 1, 1, 1)$ . By the construction of  $\tilde{\lambda}$ , it is immediate that:

$$\tilde{\lambda}_i > \tilde{\lambda}_j \Leftrightarrow \frac{\bar{\lambda}_i}{t_i} > \frac{\bar{\lambda}_j}{t_j}$$

Also by Lemma A9, we know that equilibrium access to rides in each region  $i$  is the same under  $(\tilde{\lambda}, 1, 1, 1, 1)$  and  $(\bar{\lambda}, t, \bar{c}, \alpha)$ . It immediately follows that the first half of the statement holds under  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ .

For the second half of the proposition that speaks to the regions without supply, by assumption, we know there is a  $\mu$  such that:

$$\forall i : \frac{\tilde{\lambda}_i}{1} < \mu \Leftrightarrow \tilde{n}_i = 0$$

By construction, we know  $\tilde{\lambda}_i < \mu$  is equivalent to  $\frac{\tilde{\lambda}_i}{t_i} < \alpha\beta\bar{c}\mu$ . Also, Lemma A8,  $\tilde{n}_i = 0$  is equivalent to  $\frac{1}{\beta t_i} n_i^* = 0 \Leftrightarrow n_i^* = 0$ . Therefore, we have:

$$\forall i : \frac{\tilde{\lambda}_i}{t_i} < \alpha\beta\bar{c}\mu \Leftrightarrow n_i^* = 0$$

This finishes the proof of Proposition 1 under the generic primitives in the model setting with demand-side EOD.

The logic for the proof that Proposition 2 and Proposition 3 hold under generic case if they hold for  $(\tilde{\lambda}, 1, 1, 1, 1)$  is identical to that in the proof of Lemma A5. As such, we skip the rest of the proof.

The proof of Lemma A10 is now complete.  $\square$

The last step in proving Proposition 4 is to show that it does indeed hold under  $(\tilde{\lambda}, 1, 1, 1, 1)$ . As the proof of Lemma A6 for the model with supply-side EOD, this last step is done fully numerically. The following lemma takes on this task.

**Lemma A11.** *Proposition 4 holds if  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  takes the specific form of  $(\tilde{\lambda}, 1, 1, 1, 1)$ .*

*Proof of Lemma A10.* We start by proving the second part of Proposition 1 in the context of model with demand-side EOD, and then move to all others.

*Second half of Proposition 1.* Here, we are looking for conditions on potential demand  $\bar{\lambda}_i$  such that the platform can get a positive number of drivers into region  $i$  without sustaining any loss. Note that from eq. (19), we know the region can get a positive number of drivers if:

$$\frac{c_i}{\bar{c}} \bar{\lambda}_i (1 - \alpha p_i) > 4\beta t_i$$

Plugging in  $\bar{c} = t_i = \alpha = \beta = 1$  and  $\bar{\lambda}_i = \tilde{\lambda}_i$ , this turns into:

$$c_i \tilde{\lambda}_i (1 - p_i) \geq 4$$

Or, equivalently,

$$\tilde{\lambda}_i \geq \frac{4}{c_i(1 - p_i)}$$

Then the minimum requirement for  $\tilde{\lambda}_i$  depends on the maximum possible value for  $c_i(1 - p_i)$  subject to the constraint that  $p_i \geq c_i$ . This happens when  $p_i = c_i$ . Therefore, the region will get a positive number of drivers if and only if

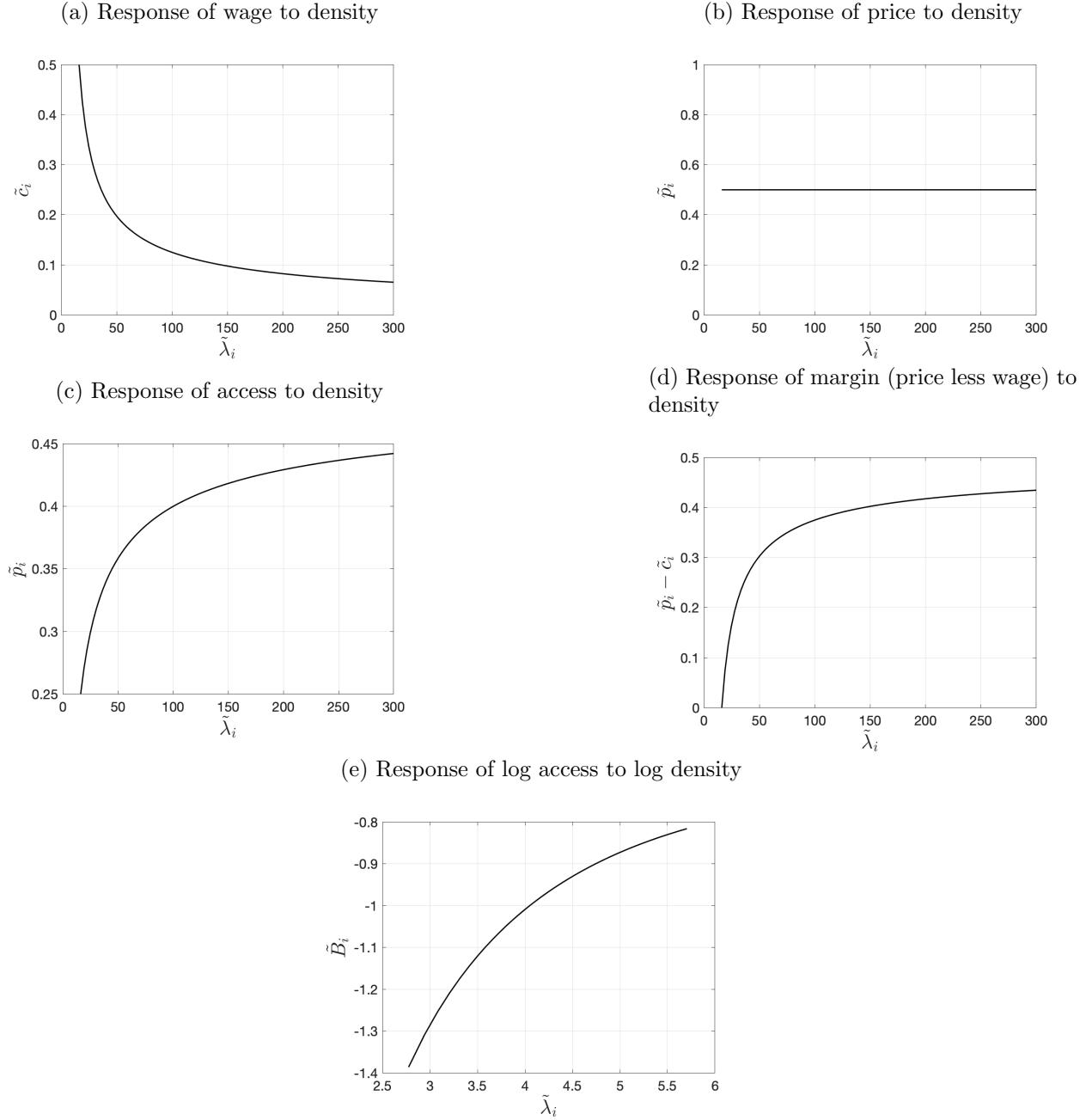
$$\tilde{\lambda}_i \geq \frac{4}{\max_{c_i} c_i(1 - c_i)} = 16$$

This proves the second part of Proposition 1 for the model with demand-side EOD.

The proof of the existence of the equilibrium in Lemma 2 is also analogous to the proof in the model with supply-side EOD. So we skip the details here. Now, we prove Lemma 2, Proposition 1 and Proposition 3 numerically with the model with demand-side EOD. To do this, we need to study the equilibrium price  $\tilde{p}_i$ , equilibrium wage  $\tilde{c}_i$ , equilibrium number of drivers  $\tilde{n}_i$ , and equilibrium access  $\tilde{A}_i$  in a given region  $i$  as a function of  $\tilde{\lambda}_i$ . Similar to the proof of Lemma A6, the key is that all of the other parameters are equal to 1 and  $\tilde{\lambda}_i$  is the only parameter changing. Therefore, for any value that  $\tilde{\lambda}_i$  assumes, we have a fully numerical problem that can be solved using a software such as R. Fig. 10 provides graphs of these equilibrium quantities as functions of  $\tilde{\lambda}_i$ .

As shown in the figure, unique optimal price and wage are computed for  $\tilde{\lambda}_i \geq 16$ , which shows that Lemma 2 holds. Second, access  $\tilde{A}_i$  is strictly increasing in density. This means if two regions  $i, j$  are such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a lower access to rides in the equilibrium. This, together with the proof of the second half of Proposition 1 given above, completes the proof of Proposition 1. Additionally, optimal wage  $\tilde{c}_i$  strictly decreases in density  $\tilde{\lambda}_i$ , while optimal price  $\tilde{p}_i$  remains constant. This means if two regions  $i, j$  are such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a higher wage in the equilibrium. Also margin  $\tilde{p}_i - \tilde{c}_i$  is strictly increasing in density. This means if two regions  $i, j$  are such that  $\tilde{\lambda}_i > \tilde{\lambda}_j$ , then region  $j$  will have a lower margin in the equilibrium. These shows that Proposition 3 holds with an exception that the optimal price  $\tilde{p}_i$  does not change with  $\tilde{\lambda}_i$ .

Figure 10: Numerical results that are necessary to prove Lemma A11.



Before showing that Proposition 2 holds with the model with demand-side EOD, we first formalize the result of constant optimal price under  $(\tilde{\lambda}, 1, 1, 1, 1)$  in the lemma below and provide the proof of it.

**Lemma A12.** *Under the theory model with demand-side EOD, for any generic primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$ , if the region  $i$  has a positive number of drivers under optimal price and wage, then the optimal price for the region is  $\frac{1}{2\alpha}$ .*

*Proof of Lemma A12.* From eq. (19), given the primitives  $(\bar{\lambda}, t, \bar{c}, \alpha, \beta)$  and  $(p, c)$ , the equilibrium

number of drivers for a region can be written as

$$n(p, c | \bar{\lambda}, t, \bar{c}, \alpha, \beta) = \frac{k + \sqrt{k^2 - 4\beta tk}}{2}, \text{ where } k = \frac{c}{\bar{c}}\bar{\lambda}(1 - \alpha p)$$

To maximize the profit, the problem that the platform solves is the following:

$$(c^*, p^*) = \arg \max_{c,p} \pi(c, p) := (p - c) \times \frac{n^*(p, c)}{c}$$

Take the first-order conditions:

$$\frac{\partial \pi}{\partial p} = \frac{n^*}{c/\bar{c}} + (p - c) \frac{1}{c/\bar{c}} \frac{\partial n^*}{\partial p} = 0 \quad (20)$$

$$\frac{\partial \pi}{\partial c} = -\frac{n^*}{c/\bar{c}} + (p - c) \frac{\frac{\partial n^*}{\partial c} c/\bar{c} - n^* \frac{1}{\bar{c}}}{(c/\bar{c})^2} = 0 \quad (21)$$

From eq. (20), we have

$$n - (p - c) \frac{\partial n^*}{\partial k} \frac{c}{\bar{c}} \bar{\lambda} \alpha = 0 \quad (22)$$

From eq. (21), we have

$$\begin{aligned} -n^* c + (p - c) \left( \frac{\partial n^*}{\partial c} c - n^* \right) &= 0 \\ -n^* p + (p - c) \frac{\partial n^*}{\partial k} \frac{c}{\bar{c}} \bar{\lambda} (1 - \alpha p) &= 0 \end{aligned} \quad (23)$$

Adding  $(1 - \alpha p)$  times eq. (22) and  $\alpha$  times eq. (23), we have

$$(1 - \alpha p)n^* - \alpha p n^* = 0$$

$$p = \frac{1}{2\alpha}$$

This finished the proof of the lemma.  $\square$

Next we prove that Proposition 2 holds under the model with demand-side EOD. Similar to the model with supply-side EOD, statement 1 follows directly from the second half of Proposition 1. As for statement 2, the second part of this statement (i.e.,  $\tilde{\lambda}_i > \tilde{\lambda}_j \Rightarrow \frac{A_j(n_j^{*'})}{A_i(n_i^{*'})} < 1$ ) is implied by Proposition 1. Then applying Lemma A7, the first part of statement 2 on the role of market thickness holds if  $B(\tau) = \log(A(e^\tau))$  is a strictly concave function. And this is shown in the last plot of Fig. 10.

To see statement 3, from Lemma A12, we have  $p^* = \frac{1}{2}$ . Then the equilibrium number of drivers of region  $i$  is given by

$$\begin{aligned} \frac{n_i}{\tilde{\lambda}_i(1 - \frac{1}{2})(1 - \frac{1}{n_i})} &= c_i \\ n_i^* &= \frac{\frac{c_i \tilde{\lambda}_i}{2} + \sqrt{\frac{c_i^2 \tilde{\lambda}_i^2}{4} - 2c_i \tilde{\lambda}_i}}{2} \end{aligned}$$

As  $\tilde{\lambda}_i$  gets larger,  $n_i^*$  gets larger. Also note that  $\tilde{\lambda}_i$  and  $n_i^*$  have the same order. So for any given  $c_i$ , the equilibrium number of rides  $r_i = \frac{n_i}{\frac{n_i}{\tilde{\lambda}_i} + \frac{t_i}{n_i}}$  approaches infinity as the market gets larger. Therefore, the optimal wage approaches 0. The access ratio of any region approaches to 1.

This finishes the proof of *Proposition 4*. ■

### E.3 Proof of Proposition 5

First suppose  $n^*$  is *not* a driver equilibrium in the version of the problem with no inter-region rides. This, by definition, means that either a small number  $\delta_0$  of new drivers can join one of the regions  $i$  from outside and strictly improve their payoffs, or a small  $\delta_0$  can relocate from  $i$  to  $i'$  or to the outside option and have a strictly positive improvement. In that case, one can find the corresponding small perturbation in  $\rho$  that would in a given amount of time, change  $n^*$  in a way that would increase the payoff in those regions. Similarly, if we assume  $(n^*, \rho^*)$  with (for a  $\rho^*$  that guarantees stability) is not an equilibrium of the version of the problem with cross-region rides, one can use the small profitable perturbation in  $\rho^*$  and use them to construct changes in  $n^*$  that yield strictly improved payoffs in some regions in the original version. This finishes the proof of the proposition. ■

## F Proofs of Propositions in the Empirical Sections

This appendix supplies proofs relevant in the empirical sections. Appendix F.1 gives the proof of Proposition 6. Appendix F.2 formalizes the statement of the identification results for the model with supply-side EOD, followed by the proof of it. Appendix F.3 formalizes the identification results for the model with demand-side EOD.

### F.1 Proof of Proposition 6

This section gives the proof of the Proposition 6 that states the relationship between access ratio and relative outflow. The proof is straightforward and follows directly from the assumptions.

*Proof of Proposition 6.* Let  $(n, \rho)$  be the steady-state driver allocation.

$$\frac{A_i(n)}{A_j(n)} = \frac{r_i/\bar{\lambda}_i}{r_j/\bar{\lambda}_j} = \frac{r_{ij}/\bar{\lambda}_{ij}}{r_{ji}/\bar{\lambda}_{ji}} = \frac{r_{ij}}{r_{ji}}$$

The second equality holds because  $\frac{r_{ij}}{\bar{\lambda}_{ij}} = \frac{r_i}{\bar{\lambda}_i}$  and  $\frac{r_{ji}}{\bar{\lambda}_{ji}} = \frac{r_j}{\bar{\lambda}_j}$ . The third equality holds because  $\bar{\lambda}_{ij} = \bar{\lambda}_{ji}$ . ■

### F.2 Proof of identification of the model with supply-side EOD

This appendix gives a formal statement of the condition of the identification of the parameters in the empirical model with supply-side EOD, followed by the proof. The condition relies on specific data patterns, which are satisfied in the data used for analysis.

First, following the discussion of the identification in the main text, recall that each  $\bar{\lambda}_{id}$  can be written as an explicit function of  $\mathcal{D}$ ,  $a$  and  $\alpha$  as in eq. (9). Therefore, the identification problem reduces to identify  $\alpha$  and  $a$  from the RO moments. We consider the relative outflow across regions for a single day  $d$ . For notational brevity, we drop the day subscript  $d$ . The relative outflow derived from the model can be written as

$$RO_{ij}^{model} = \frac{A_i}{A_j} = \frac{r_i/\bar{\lambda}_i}{r_j/\bar{\lambda}_j} = \frac{\frac{n_i}{c_i}/\bar{\lambda}_i}{\frac{n_j}{c_j}/\bar{\lambda}_j} = \frac{(1 - \alpha p_i)(1 - a \frac{s_i \bar{c}}{n_i c_i})}{(1 - \alpha p_j)(1 - a \frac{s_j \bar{c}}{n_j c_j})} \quad (24)$$

where the first equality holds following Proposition 6 and the second equality holds given the definition of access. The third equality holds given the assumption of driver equilibrium distribution. The fourth equality is obtained if we substitute  $\bar{\lambda}_i$  and  $\bar{\lambda}_j$  with  $a$ ,  $\alpha$  and observable data  $\mathcal{D}$  following eq. (9).

To uniquely identify the two parameters  $(\alpha, a)$ , it suffices to match two relative outflows between the data and the model. Therefore, the identification proof is reduced to showing that there exists two pairs of regions,  $(i, j)$  and  $(k, l)$ ,<sup>31</sup> such that the following system of equations has a unique solution for  $(a, \alpha)$ , given everything else is data:

$$\begin{aligned} RO_{ij}^{model} &= \frac{(1 - \alpha p_i)(1 - a \frac{s_i \bar{c}}{n_i c_i})}{(1 - \alpha p_j)(1 - a \frac{s_j \bar{c}}{n_j c_j})} = RO_{ij}^{data} \\ RO_{kl}^{model} &= \frac{(1 - \alpha p_k)(1 - a \frac{s_k \bar{c}}{n_k c_k})}{(1 - \alpha p_l)(1 - a \frac{s_l \bar{c}}{n_l c_l})} = RO_{kl}^{data} \end{aligned}$$

The proposition below states the condition for the uniqueness of the solution and thus the identification:

**Proposition 7.** *Under the model with supply-side economies of density, the parameters are uniquely identified if the following condition holds:*

Denote  $d_i = \frac{s_i \bar{c}}{n_i c_i}$ . There exists two distinct pairs of regions, indexed by  $(i, j)$  and  $(k, l)$ , where  $k$  may overlap with  $i$  or  $l$  may overlap with  $j$ , s.t.  $(p_i - p_j)(d_i - d_j) > 0$  and  $(p_k - p_l)(d_k - d_l) < 0$ .

Equivalently, the condition implies that, among the two pairs of regions, one pair has the same ordering for  $(p_i, p_j)$  and  $(d_i, d_j)$ , while the other pair has the opposite ordering. The implication is in line with the intuition of independence between regional price variation and regional density variation that we show in the main text. We show the formal proof below.

*Proof of Proposition 7.* We show the proof of the uniqueness of the solution in the following steps:

**Step 0:** Note that given the actual demand,  $\lambda_i = \bar{\lambda}_i(1 - \alpha p_i)$  is positive for any region  $i$ , the parameter  $\alpha$  satisfies  $\alpha < \min_{i \in I} \{ \frac{1}{p_i} \}$ . Also, given that, in equilibrium, the idle time can be

---

<sup>31</sup>Note that  $k$  or  $l$  may overlap with  $i$  or  $j$ . So we may need only three regions.

expressed as  $\frac{n_i}{\lambda_i} = \frac{c_i}{\bar{c}} - \frac{a s_i}{n_i}$ , the positive idle time implies that  $a < \min_{i \in I} \left\{ \frac{c_i n_i}{\bar{c} s_i} \right\}$ . We denote  $d_i = \frac{s_i \bar{c}}{n_i c_i}$ . Then we have  $a < \min_{i \in I} \left\{ \frac{1}{d_i} \right\}$ .

**Step 1:** Given that the condition, without loss of generality, we assume that  $p_i < p_j$ ,  $d_i < d_j$ ,  $p_k < p_l$ , and  $d_k > d_l$ .

**Step 2:** Consider a function  $f(x; k_1, k_2) = \frac{1-xk_1}{1-xk_2}$ , where  $0 < x < \min\{\frac{1}{k_1}, \frac{1}{k_2}\}$ . Then  $\frac{\partial f(x)}{\partial x} > 0$  if  $k_1 < k_2$  and  $\frac{\partial f(x)}{\partial x} < 0$  if  $k_1 > k_2$ .

Denote  $D_{ij} = \{p_i, p_j, d_i, d_j\}$  and  $D_{kl} = \{p_k, p_l, d_k, d_l\}$ . Then we can express  $RO_{ij}^{model}$  and  $RO_{kl}^{model}$  as functions of  $(a, \alpha)$ :

$$RO_{ij}^{model}(a, \alpha; D_{ij}) = \frac{(1 - \alpha p_i)(1 - ad_i)}{(1 - \alpha p_j)(1 - ad_j)} = f(\alpha; p_i, p_j) f(a; d_i, d_j)$$

$$RO_{kl}^{model}(a, \alpha; D_{kl}) = \frac{(1 - \alpha p_k)(1 - ad_k)}{(1 - \alpha p_l)(1 - ad_l)} = f(\alpha; p_k, p_l) f(a; d_k, d_l)$$

Given  $\alpha < \min_{i \in I} \left\{ \frac{1}{p_i} \right\}$ ,  $a < \min_{i \in I} \left\{ \frac{1}{d_i} \right\}$ ,  $p_i < p_j$ ,  $d_i < d_j$ ,  $p_k < p_l$ , and  $d_k > d_l$ , we have

$$\begin{aligned} \frac{\partial RO_{ij}^{model}(a, \alpha; D_{ij})}{\partial \alpha} &> 0, \quad \frac{\partial RO_{kl}^{model}(a, \alpha; D_{kl})}{\partial \alpha} > 0 \\ \frac{\partial RO_{ij}^{model}(a, \alpha; D_{ij})}{\partial a} &> 0, \quad \frac{\partial RO_{kl}^{model}(a, \alpha; D_{kl})}{\partial a} < 0 \end{aligned} \tag{25}$$

**Step 3:** Suppose there exists a pair of  $(a, \alpha)$  that satisfies the system of equations above. That is,

$$RO_{ij}^{model}(a, \alpha; D_{ij}) = RO_{ij}^{data}$$

$$RO_{kl}^{model}(a, \alpha; D_{kl}) = RO_{kl}^{data}$$

We show by contradiction that there does not exist a distinct pair of  $(a', \alpha')$  that satisfies the system of equations above, where at least one of  $a = a'$  and  $\alpha = \alpha'$  hold.

We consider the four cases below:

- (1) Consider  $a' > a$  and  $\alpha' \geq \alpha$ . Following the derivatives in eq. (25), we have  $RO_{ij}^{model}(a', \alpha'; D_{ij}) > RO_{ij}^{model}(a, \alpha; D_{ij}) = RO_{ij}^{data}$ . So  $(a', \alpha')$  cannot be the solution.
- (2) Consider  $a' > a$  and  $\alpha' < \alpha$ . Following the derivatives in eq. (25), we have  $RO_{kl}^{model}(a', \alpha'; D_{kl}) > RO_{kl}^{model}(a, \alpha; D_{kl}) = RO_{kl}^{data}$ . So  $(a', \alpha')$  cannot be the solution.
- (3) Consider  $a' < a$  and  $\alpha' \geq \alpha$ . Following the derivatives in eq. (25), we have  $RO_{kl}^{model}(a', \alpha'; D_{kl}) < RO_{kl}^{model}(a, \alpha; D_{kl}) = RO_{kl}^{data}$ . So  $(a', \alpha')$  cannot be the solution.
- (4) Consider  $a' < a$  and  $\alpha' < \alpha$ . Following the derivatives in eq. (25), we have  $RO_{ij}^{model}(a', \alpha'; D_{ij}) > RO_{ij}^{model}(a, \alpha; D_{ij}) = RO_{ij}^{data}$ . So  $(a', \alpha')$  cannot be the solution.

Therefore, there does not exist  $(a', \alpha')$  that solves the system of equations with at least one of  $a = a'$  and  $\alpha = \alpha'$  can not hold. ■

Given we observe the pattern in the data with  $(i, j)$  and  $(k, l)$  being Manhattan and Queens, Bronx and Manhattan respectively, the parameters in the model with supply-side EOD are identified in this case.

Two remarks are worth mentioning. First, the proof relies on monotonicity of  $f(x)$  given data. Therefore, it should be considered parametric identification. However, the same identification logic would essentially carry through for any non-parametric monotonic function. Second, in practice, we take the average of  $RO_{ij}^{model}$  and  $RO_{ij}^{data}$  across days. We do not require over-time variation to identify  $a$  and  $\alpha$ .

### F.3 Proof of identification of the model with demand-side EOD

Now we turn to the identification results for the model with demand-side EOD. We formalize the condition of the identification. The statement and the proof of the identification is analogous to those for the model with supply-side EOD. As such, we skip some of the details.

First, note that each  $\bar{\lambda}_{id}$  can be written as an explicit function of  $\mathcal{D}$ ,  $a$  and  $\alpha$  as in eq. (16). Therefore, the identification problem reduces to identify  $\alpha$  and  $\beta$  from the RO moments. Similar to the model with supply-side EOD, We consider the relative outflow across regions for a single day  $d$ . The relative outflow derived from the model can be written as

$$RO_{ij}^{model} = \frac{A_i}{A_j} = \frac{r_i/\bar{\lambda}_i}{r_j/\bar{\lambda}_j} = \frac{\frac{n_i}{c_i}/\bar{\lambda}_i}{\frac{n_j}{c_j}/\bar{\lambda}_j} = \frac{(1 - \alpha p_i)(1 - \beta \frac{s_i}{n_i})}{(1 - \alpha p_j)(1 - \beta \frac{s_i}{n_i})} \quad (26)$$

where the first equality holds following Proposition 6 and the second equality holds given the definition of access. The third equality holds given the assumption of driver equilibrium distribution. The fourth equality is obtained if we substitute  $\bar{\lambda}_i$  and  $\bar{\lambda}_j$  with  $a$ ,  $\alpha$  and observable data  $\mathcal{D}$  following eq. (16).

To uniquely identify the two parameters  $(\alpha, \beta)$ , it suffices to match two relative outflows between the data and the model. Therefore, the identification proof is reduced to showing that there exists two pairs of regions,  $(i, j)$  and  $(k, l)$ ,<sup>32</sup> such that the following system of equations has a unique solution for  $(\alpha, \beta)$ , given everything else is data:

$$\begin{aligned} RO_{ij}^{model} &= \frac{(1 - \alpha p_i)(1 - \beta \frac{s_i}{n_i})}{(1 - \alpha p_j)(1 - \beta \frac{s_i}{n_i})} = RO_{ij}^{data} \\ RO_{kl}^{model} &= \frac{(1 - \alpha p_k)(1 - \beta \frac{s_k}{n_k})}{(1 - \alpha p_l)(1 - \beta \frac{s_l}{n_l})} = RO_{kl}^{data} \end{aligned}$$

The proposition below states the condition for the uniqueness of the solution and thus the identification:

**Proposition 8.** *Under the model with demand-side economies of density, the parameters are uniquely identified if the following condition holds:*

---

<sup>32</sup>Note that  $k$  or  $l$  may overlap with  $i$  or  $j$ . So we may need only three regions.

Denote  $d_i = \frac{s_i}{n_i}$ . There exists two distinct pairs of regions, indexed by  $(i, j)$  and  $(k, l)$ , where  $k$  may overlap with  $i$  or  $l$  may overlap with  $j$ , s.t.  $(p_i - p_j)(d_i - d_j) > 0$  and  $(p_k - p_l)(d_k - d_l) < 0$ .

The proof is analogous to that of Proposition 7. Thus, we skip the proof. Taking to the data, the condition is satisfied if we set  $(i, j)$  to be Manhattan and Queens and  $(k, l)$  to be Bronx and Queens.

## G Empirical and Counterfactual Results for All Models

This appendix presents the complete estimation and counterfactual results for all models specified in Section 5: supply-side EOD model, demand-side EOD model, and the comprehensive model with  $\frac{\alpha}{\beta} \in \{10, 1, 0.1\}$ . Table 7 displays the average potential demand density and accesses across days in each region from each of the models. The demand densities and accesses follow the same ordering across the models with varying relative strength of supply-side and demand-side EOD.

Table 7: Model Outcomes (Corresponding to Table 5 of the paper)

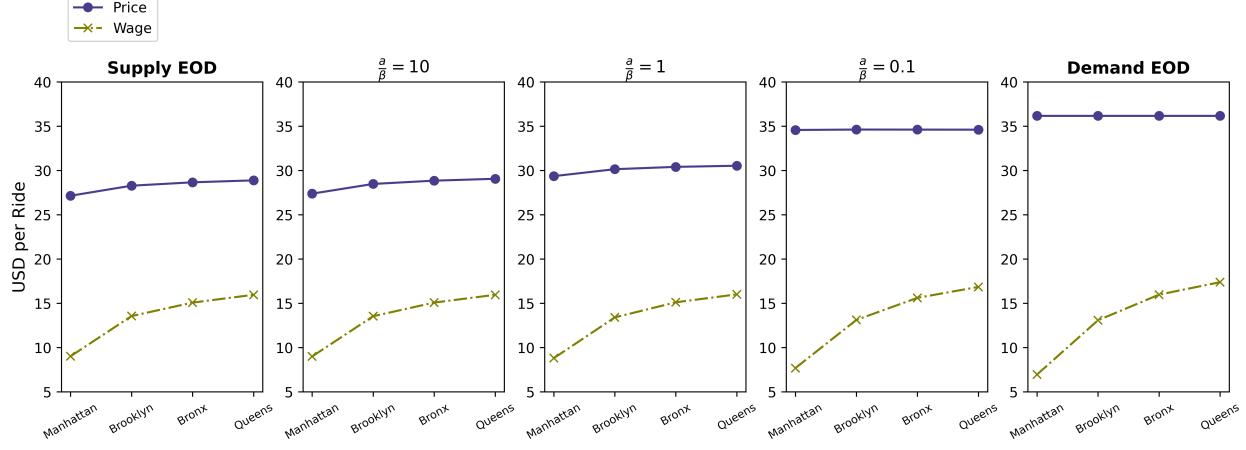
Model Spec. $\frac{\alpha}{\beta}$	Supply EOD $\infty$	Supply & Demand EOD 10	Supply & Demand EOD 1	Supply & Demand EOD 0.1	Demand EOD 0
<i>Demand Density</i> ( $\frac{\bar{\lambda}_i}{s_i}$ )					
Queens	630	627	606	554	540
Bronx	719	716	687	623	607
Brooklyn	939	936	907	835	815
Manhattan	2655	2638	2512	2272	2211
<i>Access</i> ( $\frac{r_i}{\lambda_i}$ )					
Queens	0.49	0.49	0.51	0.56	0.57
Bronx	0.51	0.52	0.54	0.59	0.61
Brooklyn	0.57	0.57	0.59	0.64	0.66
Manhattan	0.58	0.58	0.61	0.68	0.69

Note: The results presented correspond to averages across days over the data time periods. The regions are ordered by demand density.

Next, we turn to the counterfactual results. For each figure in Section 6, we provide the corresponding figure derived from the results of all five models. The implementation details and figure interpretations are similar to those described in Section 6, and therefore, are omitted here for brevity. These results demonstrate that our insights remain both qualitatively and quantitatively consistent, regardless of the source of EOD or the relative strength of supply-side versus demand-side EOD.

Figure 11: Platform optimal strategy results for all models (Corresponding to Section 6.1 of the paper)

(a) Prices and wages under platform optimal strategy with heterogeneous prices and wages



(b) Accesses with homogenous vs heterogeneous prices and wages.

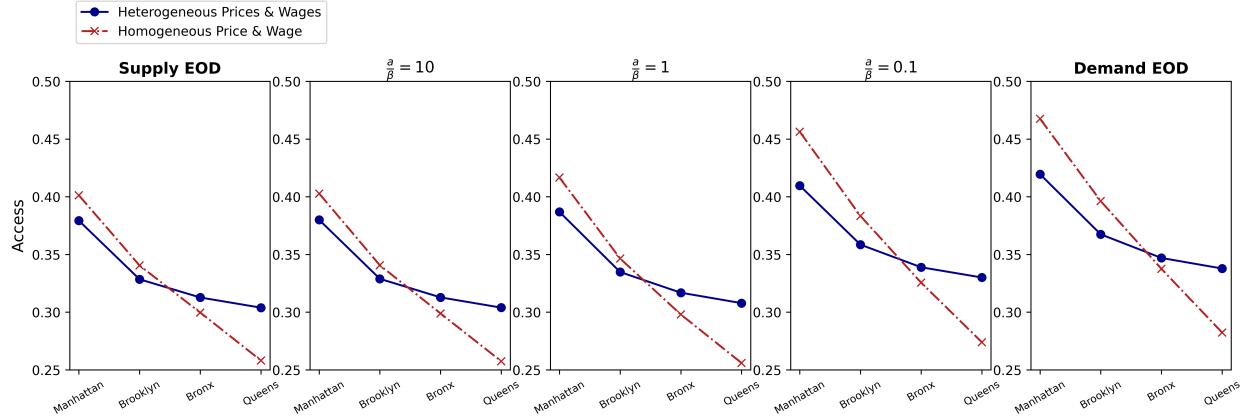


Figure 12: Access Ratio (relative to Manhattan) as platform size increases (corresponding to Section 6.2 of the paper)

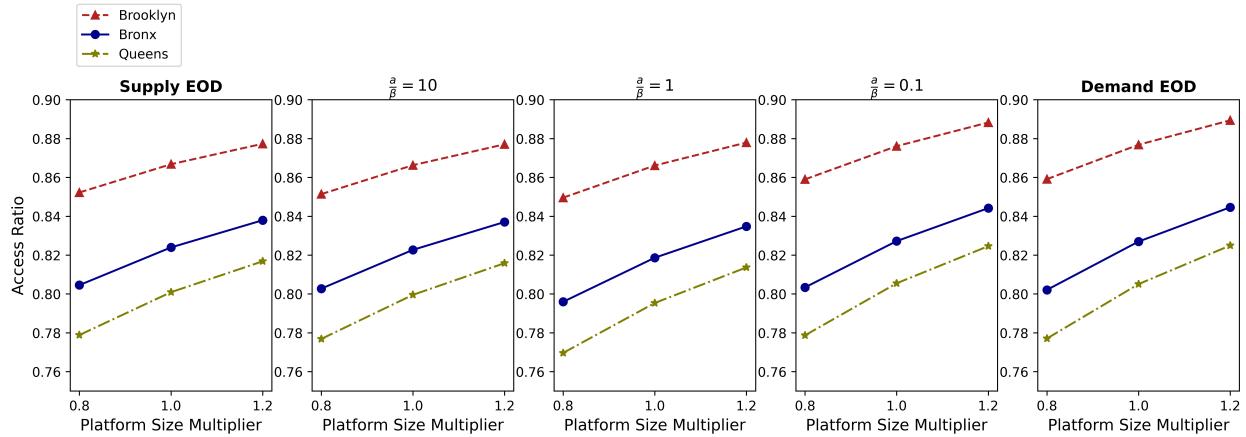
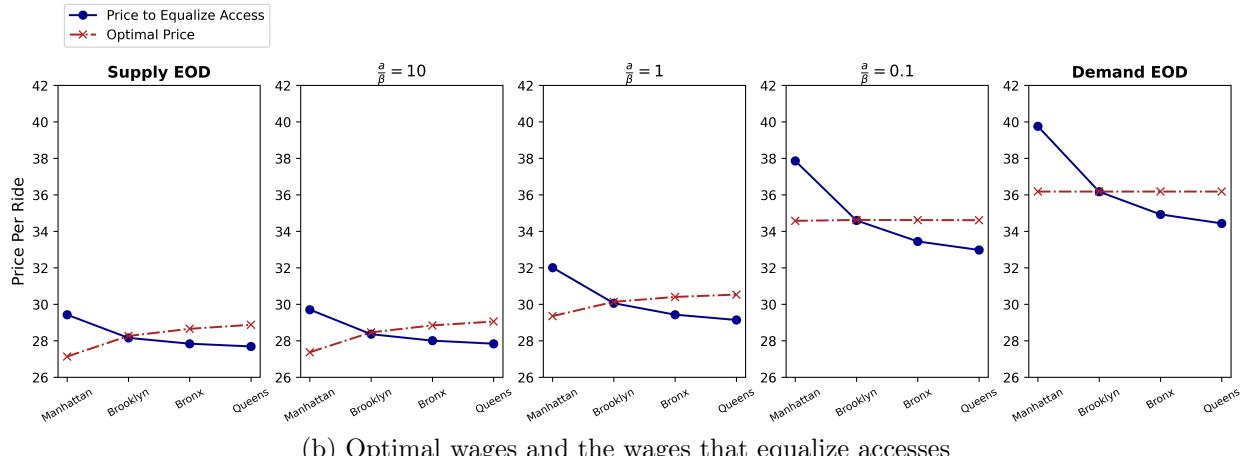
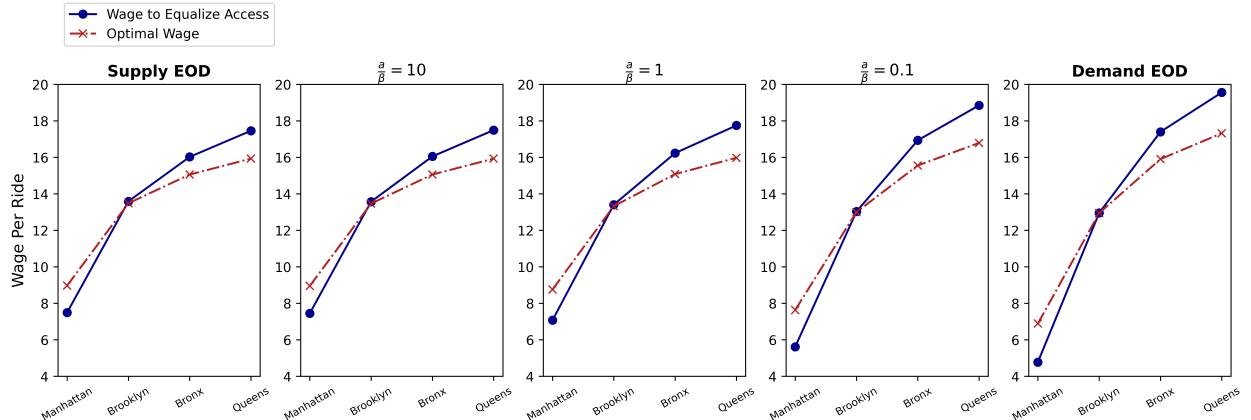


Figure 13: Equalizing accesses results for all models (corresponding to Section 6.3 of the paper).

(a) Optimal prices and the prices that equalize accesses



(b) Optimal wages and the wages that equalize accesses



## H Robustness Check: Incorporating Trip Characteristics

This appendix addresses the potential concern that the price and wage variation in the data adopted in the main analysis could be mechanically arising from the variation in trip characteristics, such as time and distance. To this end, we adapt the model to account for both trip time and trip distance. Appendix H.1 describes how we incorporate the trip characteristics into the empirical models with supply-side and demand-side EOD. Appendix H.2 presents the corresponding model estimates and counterfactual results. Our three main insights continue to hold after incorporating trip characteristics: (1) access is skewed away from less dense regions, (2) the skew is worse for smaller platforms, and (3) the platform only partially mitigates the skew.

### H.1 Incorporating Trip Characteristics

We incorporate trip characteristics into our analysis through a two-step process. First, we calculate adjusted prices and wages for each region and each day, representing values corresponding to an “average” trip characterized by the median trip time and median trip distance. Second, we include a homogeneous trip time across regions and days into the total wait time for the empirical models with supply-side and demand-side EOD. Treating this additional trip time as exogenous, we proceed with model estimation and counterfactual analyses.

#### H.1.1 Adjusted Prices and Wages

In this section, we outline the process for constructing adjusted prices and wages based on trip time and trip distance. The objective is to standardize prices and wages to reflect those corresponding to an “average” trip for each region and day. We first compute the median trip time and distance for each borough  $i$  and each day  $d$ , denoted as  $TripTime_{id}$  and  $TripDistance_{id}$ . Then we run the following regressions of price/wage on the trip time and the distance at borough-day level:

$$p_{id} = \beta_1^p TripTime_{id} + \beta_2^p TripDistance_{id} + \epsilon_{id}^p$$

$$c_{id} = \beta_1^c TripTime_{id} + \beta_2^c TripDistance_{id} + \epsilon_{id}^c$$

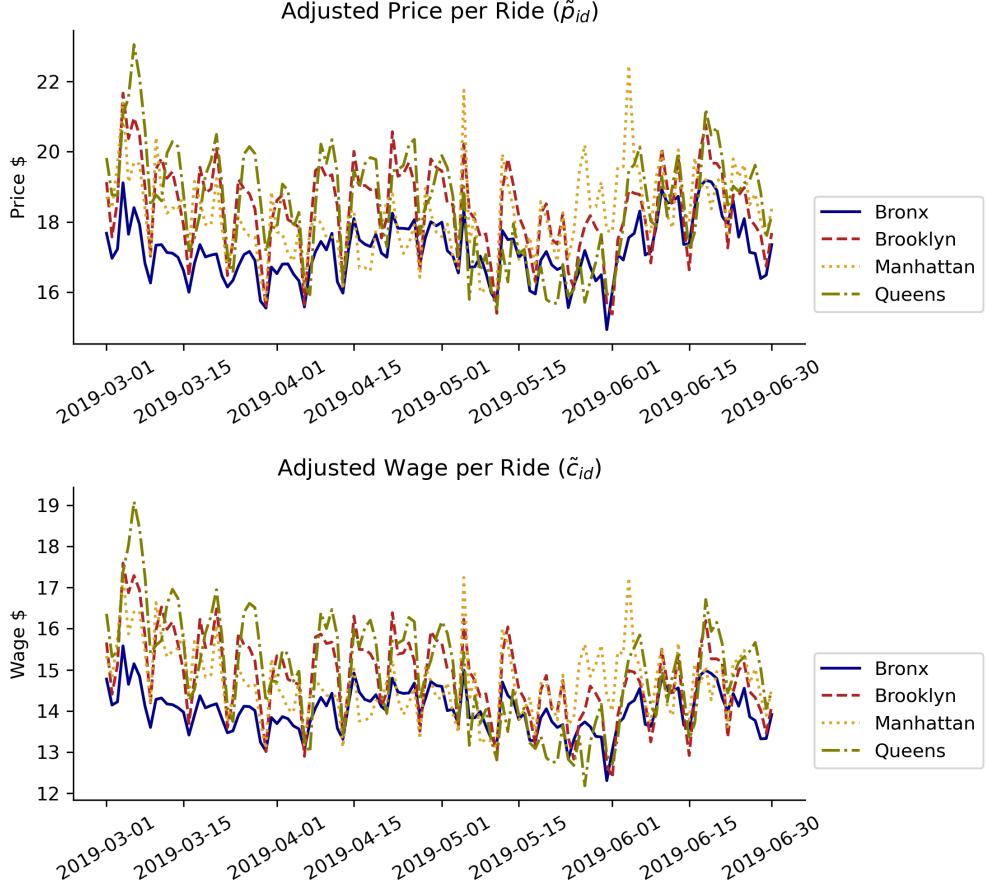
Next, we take the median trip time and trip distance across all rides across all four boroughs from March to June 2019, denoted as  $\widetilde{TripTime}$  and  $\widetilde{TripDistance}$ . Then the adjusted prices and wages can be generated as

$$\tilde{p}_{id} = \hat{\beta}_1^p \widetilde{TripTime} + \hat{\beta}_2^p \widetilde{TripDistance} + \hat{\epsilon}_{id}^p$$

$$\tilde{c}_{id} = \hat{\beta}_1^c \widetilde{TripTime} + \hat{\beta}_2^c \widetilde{TripDistance} + \hat{\epsilon}_{id}^c$$

where  $\hat{\beta}_1^p$ ,  $\hat{\beta}_2^p$ ,  $\hat{\beta}_1^c$  and  $\hat{\beta}_2^c$  are the parameter estimates, and  $\hat{\epsilon}_{id}^p$  and  $\hat{\epsilon}_{id}^c$  are the residuals obtained from the regressions. In our data, the median trip time is 14.3 minutes, and the median trip distance is 2.61 miles. Fig. 14 shows how the adjusted prices and wages vary across days and boroughs.

Figure 14: Variation of adjusted prices and wages across days and boroughs.



We use  $\tilde{p}_{id}$  and  $\tilde{c}_{id}$  for calibration instead of  $p_{id}$  and  $c_{id}$ , which are directly averaged from the data, while keeping all other data inputs unchanged. As such, variations in trip time and duration do not carry over to variations in the adjusted price and wage measures.

### H.1.2 Incorporating Trip Time to the Empirical Models

**Supply-Side EOD Model** Let  $h_{id}$  denote the trip time for region  $i$  on day  $d$ .  $h_{id} \in \mathcal{D}$ , i.e. it belongs to the set of the observable data. With trip time  $h_{id}$ , the total wait time is modeled as

$$W_{id}(n_{id}) = h_{id} + \frac{n_{id}}{\bar{\lambda}(1 - \alpha p_{id})} + \frac{as_i}{n_i}$$

The parameters to be estimated remain the same as the main model specification in the paper, namely  $(a, \alpha, \{\bar{\lambda}_{id}\}_{i \in I, d \in D})$ .

**Demand-Side EOD Model** With trip time  $h_{id}$ , the total wait time in the demand-side EOD model is given by

$$W_{id}(n_{id}) = h_{id} + \frac{n_{id}}{\bar{\lambda}(1 - \alpha p_{id})(1 - \beta \frac{s_i}{n_{id}})}$$

The parameters to be estimated are  $(a, \beta, \{\bar{\lambda}_{id}\}_{i \in I, d \in D})$ .

**Implementing Model Estimation and Counterfactual Simulations** To estimate the supply-side and demand-side EOD models with a homogeneous trip time, we utilize the standardized  $\tilde{p}_{id}$  and  $\tilde{p}_{id}$  for an “average” trip as derived in Appendix H.1.1, which replaces the raw average prices and wages,  $p_{id}$  and  $c_{id}$ , used in the paper. Accordingly, we set  $h_{id} = \tilde{h}$ , which represents the median trip time calculated across all trips within the data time span and across all regions.

Using the model estimates, we conduct the counterfactual analyses under the assumption that  $h_{id} = \tilde{h} \forall i, d$  is fixed and exogenous. Apart from incorporating trip time as part of the total wait time, all other aspects of the counterfactual analysis implementation remain the same as the approach outlined in the main paper. We omit the details in this appendix for brevity. Overall, we show in the following section that our primary insights continue to hold.

## H.2 Model Estimation and Counterfactual Results

Table 8 presents the parameter estimates obtained from the model incorporating trip time. The estimates are of a similar scale in both the supply-side and demand-side EOD models.

Table 8: Model Estimates with Trip Time

	Supply-Side EOD	Demand-Side EOD
$\alpha$	0.0312 (0.0012)	0.0282 (0.0013)
$a$	14.95 (0.53)	N/A
$\beta$	N/A	41.05 (1.35)
$\bar{\lambda}$		
Bronx	11008	9772
Brooklyn	19544	17361
Manhattan	29087	25806
Queens	15738	13960

Table 9 compares the total number of drivers per hour across boroughs between our main specification and the model with trip time. Specifically, we sum the estimates of  $n_{id}$  across regions to get the number of drivers per hour for each day:  $\tilde{n}_d = \sum_i n_{id}$ , and calculate the average of  $\tilde{n}_d$  across days for each month. The numbers are largely consistent between the two models.

Table 10 compares prices and wages under the optimal heterogeneous prices and wages with those from the main model specification. On average, with trip time included, the optimal price

Table 9: Comparison of the Total Number of Drivers per Hour Across Boroughs Between Models

Month	Model w/o Trip Time	Model with Trip Time
March	11613	11483
April	10507	10485
May	10715	10289
June	10523	10220

decreases by 30% with the supply-side model and 38% with the demand-side model. The average change in optimal wage is about 9% in supply-side model and 26% in demand-side model.

Table 10: Counterfactual optimal prices and wages

(a) Optimal Price				
	Supply-Side EOD		Demand-Side EOD	
	No Trip Time	With Trip Time	No Trip Time	With Trip Time
Manhattan	27.15	21.63	36.18	23.20
Brooklyn	28.28	21.77	36.18	22.57
Bronx	28.66	21.87	36.18	22.18
Queens	28.88	21.92	36.18	22.01

(b) Optimal Wage				
	Supply-Side EOD		Demand-Side EOD	
	No Trip Time	With Trip Time	No Trip Time	With Trip Time
Manhattan	9.01	12.46	6.95	12.15
Brooklyn	13.55	14.14	13.08	13.36
Bronx	15.08	15.02	15.97	14.29
Queens	15.96	15.39	17.38	14.75

Fig. 15 illustrates access across regions under both heterogeneous and homogeneous prices and wages. The figure reveals that access is skewed toward denser regions. Furthermore, compared to homogeneous pricing, access levels under heterogeneous pricing are distributed within a narrower range, indicating that the platform partially mitigates, though does not fully eliminate, disparities in access.

Figure 15: Region access under heterogeneous and homogeneous pricing

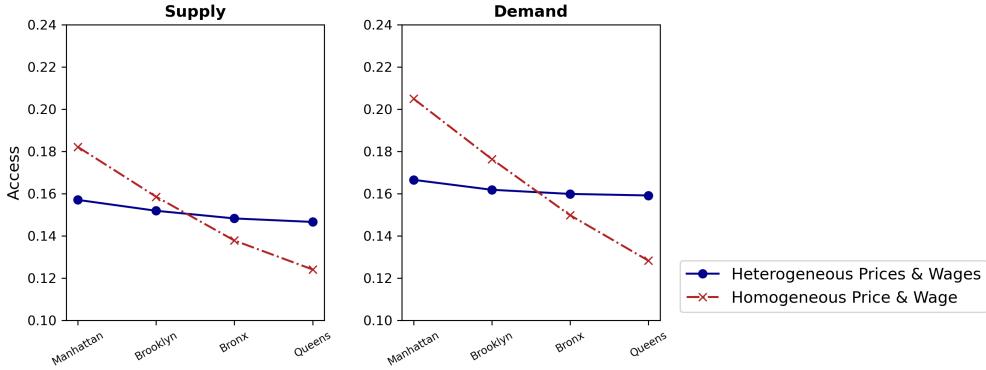
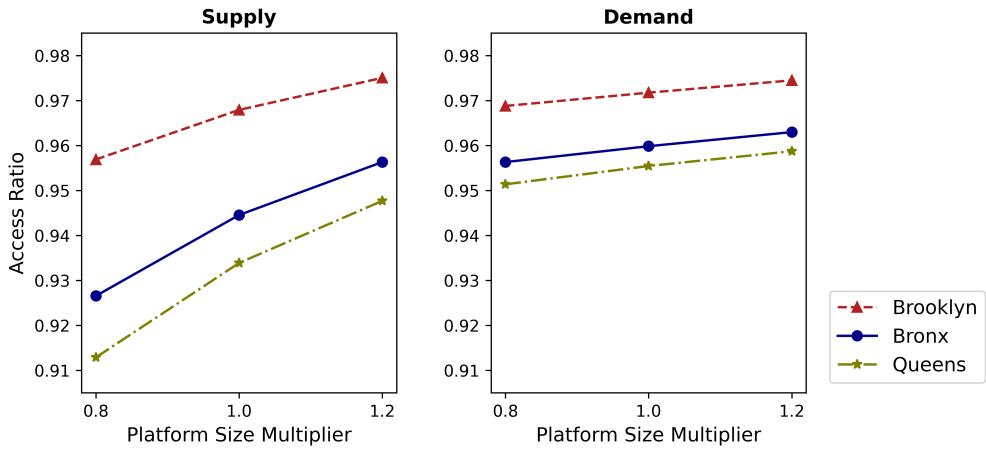


Fig. 16 illustrates access ratios relative to Manhattan (the densest region) as the platform size increases. Ratios approaching 1 indicates that access skew is mitigated as the platform grows larger. Combined with Fig. 15, the results in this section confirm that the three main insights of the paper remain robust to the inclusion of trip time: access is skewed away from less dense regions, the skew is worse for smaller platforms, and the platform only partially mitigates the skew.

Figure 16: Access ratios versus Manhattan (the densest region) with different platform sizes.



## I Price Trend after 2019

In Section 6.1, our counterfactual analysis of platform optimal pricing predicts that overall the optimal price is higher than what Uber charged in 2019. One potential explanation could be the competition between Uber and other ridesharing platforms. Accordingly, it is expected that the prices increase as the competition is softened when Uber grows larger. This section shows that our prediction is in line with what Uber has done between 2019 and 2023.<sup>33</sup>

Table 11 lists the 25th percentile, median and 75th percentile of prices among all the rides

---

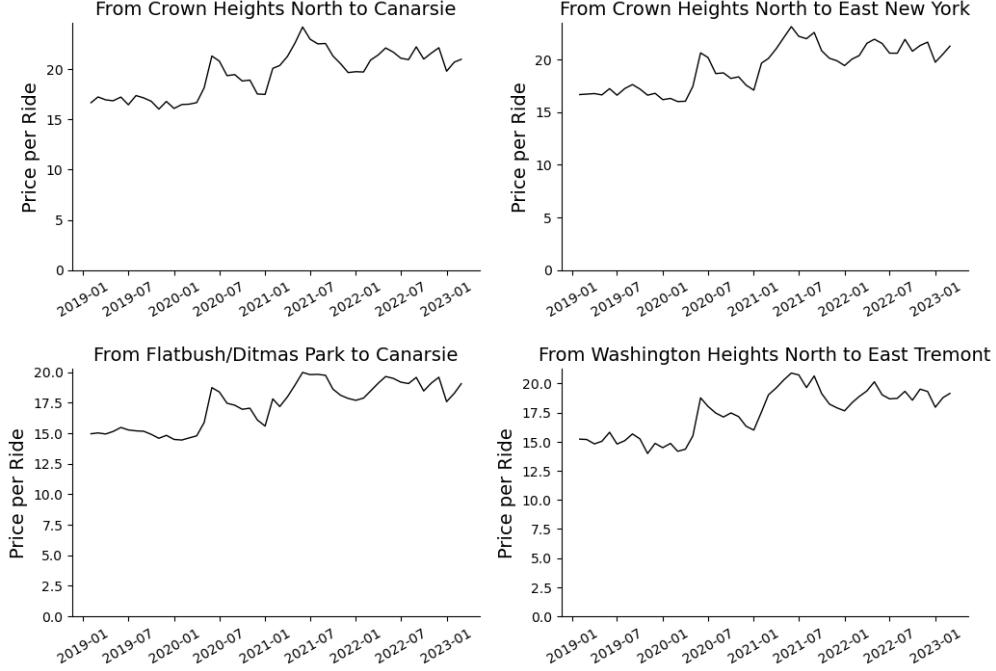
<sup>33</sup>The magnitude of price increase would be smaller if considering inflation.

taken in March 2019 and March 2023 respectively. Overall, the price has increased by about since March 2019. Fig. 17 presents the monthly average price of sampled frequently requested routes between pairs of taxi zones from the beginning of 2019 to the end of 2023. The trend agrees with the prediction of the counterfactual.

Table 11: Summary Statistics of Price per Ride in March 2019 and March 2023

Borough	25th percentile		Median		75th percentile	
	March 2019	March 2023	March 2019	March 2023	March 2019	March 2023
Bronx	13.22	17.52	13.56	17.84	13.93	18.30
Brooklyn	16.63	19.95	16.84	20.46	17.38	20.76
Manhattan	19.01	23.29	19.61	24.00	20.26	25.04
Queens	21.54	23.38	21.92	24.03	23.24	24.60

Figure 17: Average price per ride by month from 2019 to 2023 for sampled routes (between pairs of taxi zones).



## J Comparison of Data Range to Rosaia (2023)

This appendix compares the geographical locations we include in our analysis to those included in Rosaia (2023). Fig. 18 shows the map of regions included in Rosaia (2023). Fig. 19 presents the variation of net incoming trips across the locations included in our data and that across the locations included in Rosaia (2023) data. Fig. 20 compares the variation of net outgoing trips

and Fig. 21 displays the population density. For all three measures, the geographical span of the data we use encompasses additional sparser regions compared to Rosaia (2023) and thus is more amenable to capture economies of density.

Figure 18: Taxi Zones included in Rosaia (2023); copied from Rosaia (2020) Figure 1



Figure 19: Total Number of Net Incoming Trips in June 2019 by Taxi Zones

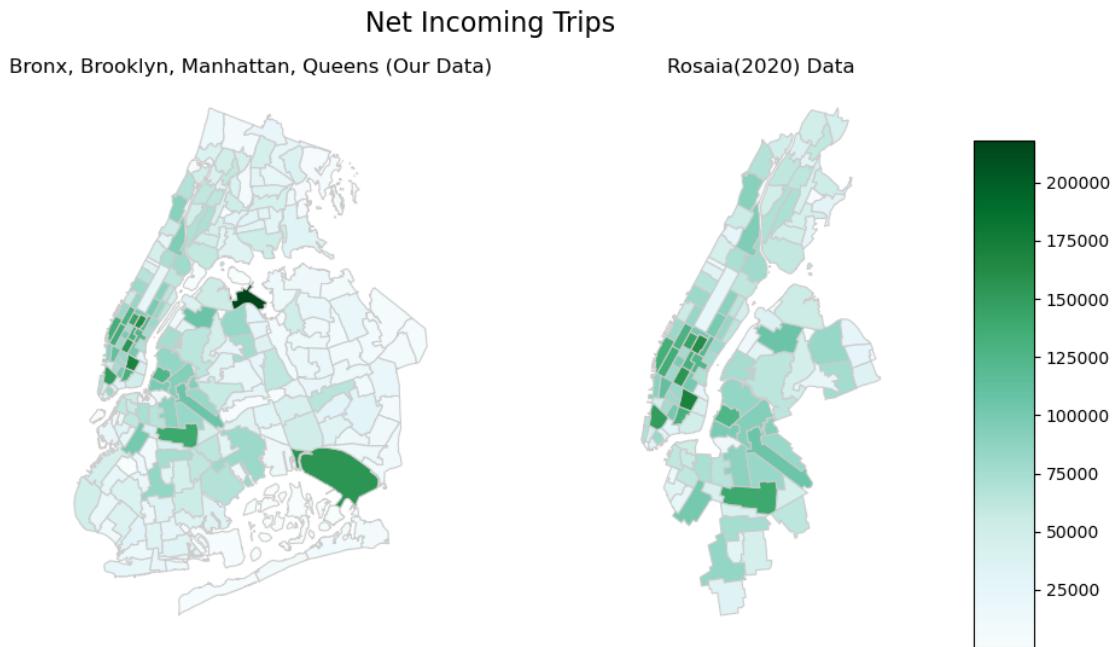


Figure 20: Total Number of Net Outgoing Trips in June 2019 by Taxi Zones

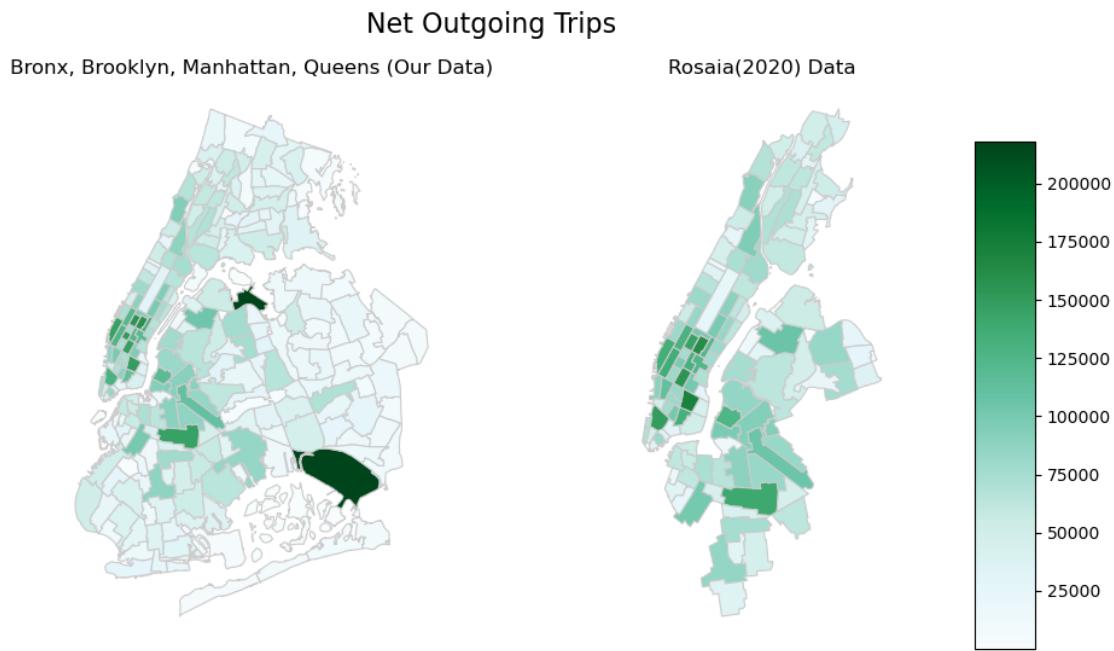
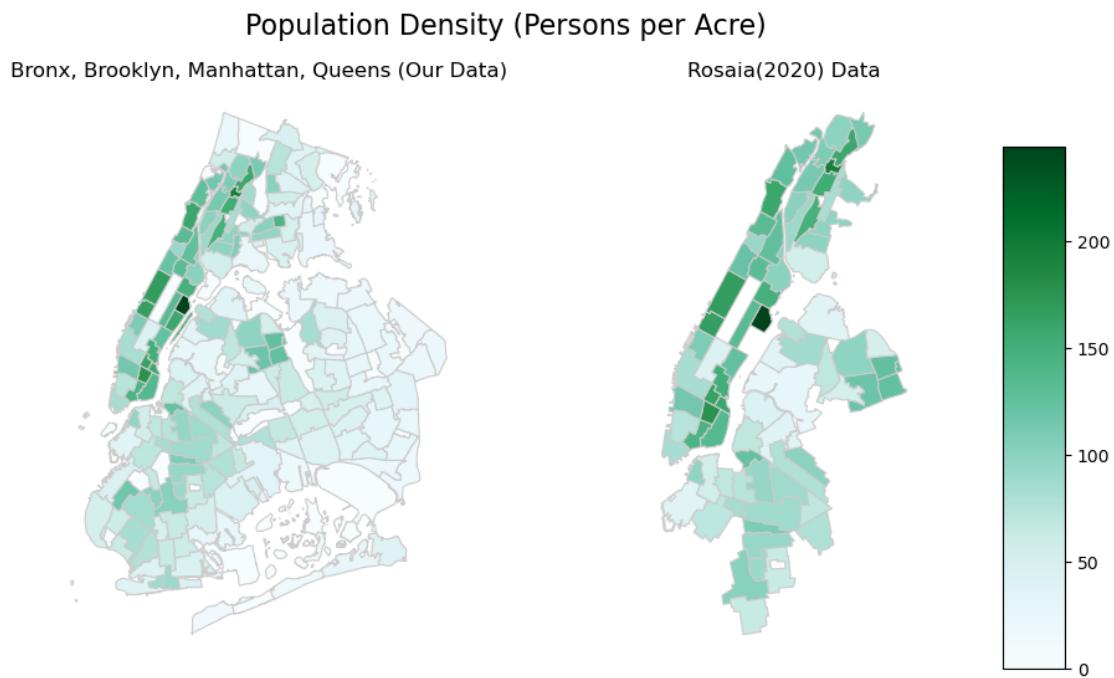


Figure 21: Population Density by Neighborhood. Note that the neighborhoods are not divided as the same as the taxi zones, so the areas on the right hand side do not exactly coincide with Rosaia's.



## References

- Frechette, G. R., Lizzeri, A., and Salz, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, 109(8):2954–2992.
- Rosaia, N. (2020). Competing platforms and transport equilibrium: Evidence from new york city.
- Rosaia, N. (2023). Competing platforms and transport equilibrium.