Incentivizing flexible workers in the gig economy: The case of ride-hailing

Cemil Selcuk (Corresponding author)
Department of Economics, Cardiff University, Cardiff, CF10 3EU, United Kingdom
E-mail: selcukc@cardiff.ac.uk

Bilal Gokpinar
UCL School of Management, University College London, London E14 5AA, United Kingdom
E-mail: b.gokpinar@ucl.ac.uk

Abstract

On-demand platforms like ride-sharing services rely heavily on economic incentives to attract, retain, and manage independent workers who have significant discretion over whether and where to work. Using an analytically tractable spatial model, we explore the impact of different pricing and commission strategies on customer demand, driver entry and retention, and their location choices. Our model yields several unique results and actionable insights. We find that flexible commission policies are more effective than fixed commission policies in allocating drivers efficiently across locations, reducing bottlenecks, and improving driver retention. We also show that commission-based interventions are more effective than price interventions in responding to labor market changes, as they directly affect driver incentives without distorting customer demand. Finally, if fairness-sensitive customers are prevalent in the market, then fixed pricing, combined with flexible commissions, becomes the optimal rule. Simulations based on actual ride patterns from New York City and Los Angeles confirm our insights.

Keywords: Ride-sharing, worker entry and retention, gig economy, driver incentives
Received: February 2023; accepted: November 2025 by Geoffrey Parker after three revisions.

1 Introduction

Uber operates in over 70 countries and 10,000 cities, and it has reported an average of 28 million trips per day globally in 2024 (Uber, 2024). Finding and keeping drivers have always been a challenge for ride-hailing platforms as they experience significant driver turnover (Brown, 2019; Cook et al., 2020). Indeed, according to a report by Uber, 11% of new drivers stop driving within a month, and about half of them leave within a year (Huet, 2015). Flexible labor supply in ride-sharing platforms has been further affected by the recent emergence of alternative work options such as food and grocery delivery (Bursztynsky, 2021), and ongoing issues with drivers' employment rights, working and pay conditions (Paul, 2021).

In modern ride-sharing, platforms engage with a large number of drivers whose participation and retention are highly responsive to earning opportunities. When a platform offers attractive earnings—through high fares and favorable commission rates—drivers would be more inclined to log in and accept rides. Conversely, if the incentives are insufficient, drivers might decide to log off and stop accepting rides. For instance, a driver might use both Uber and Lyft apps simultaneously and switch between them (Allon et al., 2023a). Additionally, it is important to recognize that these dynamics are also influenced by local labor market conditions, which can differ from city to city.

Pricing plays a crucial dual role in this ecosystem by influencing both drivers' expected earnings and customer demand. Generally, drivers prefer higher prices, but such prices can deter potential riders and decrease demand. Conversely, lower prices make rides more affordable, boosting demand. Finally, the spatial differentiation of supply and demand can create varying earning opportunities, attracting more drivers to certain locations and impacting service availability in other areas.

Given these dynamics, the platform must adopt a comprehensive approach in its operational policy decisions, considering the interplay between pricing policies, driver entry and exit, customer demand, and drivers' location choices. That is, the platform needs to create the right incentives using alternative compensation schemes in the right market, retain the right fleet size, serve the right locations, and ultimately maintain a smooth and successful operation in serving customers. To study the strategic impact of such incentives, we develop an analytically tractable model based on four key features.

First, the model is spatial in that the platform and the drivers operate over a network of locations with differing distances and traffic flows. Indeed, some locations have fundamentally different demand patterns as they consistently pull in and send away more traffic than others[1]. The spatial structure allows the platform to account for such variations in local demand and supply, as well as the self-selection of drivers who choose where to search for passengers based on location-specific factors such as local demand, price, and commission rates. By employing a spatial model, the platform effectively takes a system-wide view, choosing prices and commissions while anticipating how each decision shapes city-wide outcomes.

Second, we consider four alternative operational models: (1) Fixed price, fixed commission; (2) Flexible price, fixed commission; (3) Fixed price, flexible commission; and (4) Flexible price, flexible commission, comparing their performance in terms of the number of matches and profits. These four schemes capture a wide range of real-world practices employed by ride-sharing platforms such as Uber and Lyft. Under fixed pricing, the platform charges a uniform, per-mile rate across the entire city. In contrast, flexible pricing involves location-specific rates. Uber, for instance, employs "route-based pricing", setting different prices for routes based on "their understanding of demand patterns". This approach is justified by the fact that "traveling between a fancy neighborhood and a city center [...] might cost a premium rate" (Mahdawi, 2018).[2] This is akin to the notion of flexible pricing in our paper. In contrast, Lyft's "Price Lock" allows riders to subscribe to fixed fares for specific recurring routes, turning the typical dynamic pricing into predictable fixed prices (Morrow, 2024). Fixed and flexible commission rates are defined likewise. While the fixed commission model is more common—e.g., in most cities, Uber takes a fixed 25% commission and drivers keep 75% of the revenue (Uber, 2020)—some platforms have experimented with schemes involving location-specific or dynamic commissions. For example, Lyft has implemented "Bonus Zones," where drivers receive

[1]For instance, based on actual ride patterns, we observe that in Los Angeles there is consistently high flow of traffic to and from Santa Monica, West Hollywood and the Los Angeles International Airport, but the same is not true for, say, Studio City or Pacific Palisades.

[2]We should mention that our notion of flexible pricing, which is akin to Uber's route-based pricing, is different from "surge pricing", which addresses temporary demand spikes due to events like major sports games or bad weather. Instead, we note that some locations have fundamentally different demand patterns. Such persistent long-term differences call for flexible schemes in which prices and commission rates can be conditioned on location-specific factors.

bonus payments for picking up riders in certain areas, thereby creating location-specific incentives under fixed pricing—a structure that seems to correspond to fixed-price flexible commission model (Model 3) in our paper.[3]

Third, the platform faces a flexible labor supply, which makes driver entry endogenous in our model. By setting prices and commissions, the platform determines the expected earnings for drivers, which, in turn, determines the amount of driver entry. These policies also influence how drivers distribute themselves across locations based on earning potential. As a result, the choice of prices and commissions impacts both the size of the driver fleet as well as its allocation in the city.

Fourth, drivers' participation depends on their ability to secure customers. Those who repeatedly struggle to find customers may become "discouraged" and ultimately leave the platform (as discussed further below). That is, driver retention is also endogenous in our model. Finally, in Section 5.2, we further consider the presence of behavioral customers who may react negatively to location-specific pricing due to perceived fairness concerns.

Our analysis reveals important insights as well as several actionable operational strategies by the platform. Our first results pertains to the interplay between operational policies, driver incentives, and fleet size. A primary challenge for the platform is ensuring an even distribution of drivers across the city, particularly by incentivizing them towards undesirable locations, as failing to do so could turn such locations into bottlenecks. With fixed commission models (Operational Models 1 and 2), the platform can address this issue only through price interventions, which not only distort the demand but also do a poor job of incentivizing drivers. In contrast, flexible commission policies (Models 3 and 4) can efficiently utilize available vehicles for rides without resorting to unnecessary price hikes.

This difference has further implications when considering free entry. Due to their underutilization of available vehicles, fixed commission models often result in a larger fleet size. In contrast, flexible commission models prevent bottlenecks and maximize the use of available vehicles, thereby reducing the need for a large fleet. A smaller fleet size, however, does not equate to fewer matches or reduced profits. Indeed, our analysis demonstrates that flexible commission models, by effectively utilizing drivers, generate more matches, higher profits, and greater consumer surplus. This presents an opportunity for the platform to leverage flexible commissions which can enhance its operational performance.

Our second result highlights a key connection between flexible commissions and driver retention. Consistent with recent research on driver motivation (Hall and Krueger, 2018; Allon et al., 2023a,b), we recognize that drivers may drop out if they struggle to find passengers, while successful matches encourage continued participation. This notion yields an important insight. Under a fixed commission system, to mitigate the dropout risk, the platform must reduce prices to stimulate demand and help drivers secure matches. By contrast, with flexible commissions, such price adjustments are unnecessary. Indeed, flexible commission systems utilize the driver fleet more efficiently, so drivers consistently find customers and, therefore, are unlikely to drop out prematurely. Consequently, the platform benefits from not having to deal with the disruption associated with the frequent turnover of drivers. This represents a significant operational advantage for the platform.

Our third result focuses on how the platform optimally responds to changes in the local labor

[3]Ola, a platform from India, has piloted a flat-fee model where drivers pay a fixed daily amount to access the platform and keep all fare revenue (OutlookBusiness, 2025). Bolt, on the other hand, has tested performance-based commissions, offering reduced commission rates to top-rated drivers (AInvest, 2025). While these are not directly location-contingent in the sense modeled here, they reflect a broader trend toward flexible, nonstandard approaches to driver compensation.

market conditions. The extent of driver entry depends on the attractiveness of earning opportunities. When drivers become less responsive to such opportunities—indicating a reduced market sensitivity—the platform faces a challenge in maintaining an adequate number of drivers to meet the demand. To address this, the platform must improve expected earnings. Our analysis suggests that this is best achieved by raising commissions (drivers' share of the revenue), rather than prices. This is because higher commissions incentivize drivers without suppressing customer demand, whereas higher prices reduce demand, making them a suboptimal tool for attracting drivers. Our result, therefore, has a clear actionable insight: when faced with changing labor market conditions, the platform should prioritize the use of commissions. Doing so successfully incentivizes drivers and ensures an adequately sized driver fleet while preserving a stable demand.

Fourth, to illustrate our findings in a real-world setting, we calibrate the model for New York City and Los Angeles based on ride patterns we extracted from a publicly available connectome map on Uber's website. Our simulations reveal two important insights: (i) We document that the performance of operating models depends on how balanced a city's traffic structure is in terms of trip lengths and traffic flows. If these parameters show significant variation across the city, then pursuing a non-flexible policy is more "costly" for the platform. Because Los Angeles has a more imbalanced traffic structure than New York in our data, non-flexible rules fare worse in Los Angeles than in New York. Our subsequent simulations based on randomly generated cities with varying distances and transition matrices further confirm this insight. (ii) Since Model 4 (flexible commission, flexible price) encompasses the other operating models as special cases, it outperforms them in generating profits. Interestingly, however, the performance difference between Model 3 (flexible commission, fixed price) and Model 4 is minimal. This observation highlights the importance of flexible commissions in preventing bottlenecks and efficiently utilizing drivers across the city. Once this aspect is accounted for, the advantage of pursuing a location-specific pricing scheme seems to be relatively small.

Finally, our fifth result provides a more nuanced connection between customer behavior and the optimal operating policy. As noted above, Model 4 outperforms the other operating models and is, in principle, the platform's best option. However, the fact that Model 3—which relies on uniform pricing—performs nearly as well introduces an important caveat. Model 4 incorporates location-specific pricing, meaning that certain areas may face significantly higher prices than others. This can alienate behavioral customers who perceive such differences as unfair and respond by disengaging from the platform.[4] If the share of such customers exceeds a threshold, Model 4's profitability falls below that of Model 3. This result offers a key managerial insight for the platform. Behavioral reactions by customers can offset the benefits of location-specific pricing. When a large segment of customers responds this way, a uniform pricing strategy, combined with flexible commissions, becomes the optimal approach.

2 Related Literature

Our study focuses on ride-hailing with the objective of efficiently matching riders and drivers. To provide context, we briefly review the relevant literature on the taxi industry that forms the

[4]Though somewhat different, Uber's surge pricing resembles the pricing scheme in Model 4, and the backlash against it suggests that some customers indeed perceive such practices as unfair. It is described as "price gouging" by The New York Times (Lowrey, 2014), "exploitative" by Harvard Business Review (Dholakia, 2015), and a "scam" by CNN Business (Morrow, 2024).

foundation of our work. Lagos (2000) highlights endogenous search frictions in the taxi-cab market. Buchholz (2022) considers a non-stationary environment by employing data from New York City and analyzing the dynamic spatial equilibrium of taxi-cabs. Our research extends this body of work by incorporating a platform that sets prices and commission rates, whereas in the aforementioned models, there is no platform and prices are exogenous.

Our work is related to two-sided markets (Parker and Van Alstyne, 2005; Rochet and Tirole, 2006; Armstrong, 2006) and the literature on peer-to-peer matching platforms (Einav et al., 2016; Cramer and Krueger, 2016; Benjaafar and Hu, 2020) with a focus on ride-hailing (Wang et al., 2019; Chakravarty, 2021; Naumov and Keith, 2022). While the study of pricing strategies has a long history in the two-sided markets literature (Rochet and Tirole, 2003; Parker and Van Alstyne, 2005; Eisenmann et al., 2006; Weyl, 2010; Tan et al., 2020), there has been increased attention on the design of on-demand ride-hailing platforms and corresponding incentive schemes with the ultimate goal of better matching demand with supply (Cachon et al., 2017; Bai et al., 2019). However, most of these studies have focused on addressing short-term demand fluctuations with dynamic surge pricing (Chen and Sheldon, 2015; Banerjee et al., 2015; Castillo et al., 2017; Castillo, 2023).

Ride-sharing platforms' pricing, wage, and compensation decisions have attracted attention (Cachon et al., 2017; Hu and Zhou, 2019; Cohen and Zhang, 2022); however, these studies do not explicitly take into account spatial features of the city where the platform operates. Indeed, a key aspect of the process of matching demand with supply in ride-sharing is the spatial differentiation of consumer demand and the direct influence of pricing policies on the strategic search behavior of drivers across various locations, which has received relatively little attention in the literature. Exceptions include Guda and Subramanian (2019) who study surge pricing and information sharing in a two-zone-two-period setup and more importantly Bimpikis et al. (2019) who explore spatial price discrimination for a ride-sharing platform.

We significantly extend this line of work by concentrating on driver entry, driver retention, and customer fairness concerns—factors that increasingly influence platform operations. We build on a spatial model while differing from Bimpikis et al. (2019) in several important aspects. First, driver entry in our model is endogenous, and the optimal fleet size depends on the labor market sensitivity to earning opportunities. Second, driver retention is also endogenous, as drivers may become discouraged and drop out prematurely if they cannot find enough matches. Finally, we consider the presence of behavioral customers who may react negatively to location-specific pricing due to perceived fairness concerns. Thanks to these novel features, our model yields several unique results and actionable insights. To summarize briefly: (i) Adopting a flexible (location-specific) commission policy leads to more matches, which in turn improves driver retention and reduces operational disruptions[5]. (ii) Adjusting commissions, rather than prices, is a more effective way to respond to labor market fluctuations[6]. (iii) If fairness-sensitive customers are prevalent, then

[5]This result obtains because flexible-commission models with free entry lead to equilibrium outcomes in which all active drivers are fully utilized, so that no driver prematurely exits. The nature of this equilibrium and its impact on driver retention is not trivial and to our knowledge, has not been documented in the literature.

[6]When the labor supply becomes less responsive, the platform's profit will be negatively affected, and the platform should respond by improving the expected earnings of the drivers. However, it is not clear how the platform should adjust its policies to achieve this. Both prices and commissions can, in principle, be used to improve earnings to induce more driver entry, but their equilibrium implications are hard to predict. Our analysis—both theoretical and numerical—shows that the platform should rely primarily on commission adjustments, with price adjustments playing a peripheral role. This result is not immediate and we believe our study is the first in the literature to identify and explain this specific nature of the policy response.

uniform pricing, combined with flexible commissions, becomes the optimal rule[7]. To the best of our knowledge, these assumptions and the resulting managerial insights are unique to our model, making it a distinct and significant contribution to the literature.

Finally, our work also has broad connections with the literature on incentives and compensation plan design (Jain, 2012; Chan et al., 2014). Previous literature explored how to best align incentives of flexible workers with those of the firm by considering commissions and bonuses (Schöttner, 2017). More recent work focused on two-sided market platforms and examined the compensation of salespeople employed by such platforms in the presence of network effects (Bhargava and Rubel, 2019). A common aim of this literature is to understand how different compensation schemes affect the effort choices of salespeople who have considerable autonomy and flexibility in their work. In a similar spirit, our study investigates how an on-demand platform designs incentives to manage a highly independent and flexible workforce effectively.

3 Model

3.1 Environment

Time is discrete and continues forever. We consider a city that consists of $n \geq 2$ locations and is populated by a continuum of passengers with size 1 and a continuum of cars with size θ. The number of passengers and cars at location i are denoted by $y_i > 0$ and $x_i > 0$ and they satisfy $\sum_{i=1}^n y_i = 1$ and $\sum_{i=1}^n x_i = \theta$. The physical distance between locations i and j is denoted by $\delta_{i,j}$ and people's moves across these locations are governed by a Markov process, characterized by the exogenous row stochastic transition matrix $T = (a_{i,j})_{n\times n}$ where $a_{i,j} > 0$ denotes the probability that a person at location i wishes to travel to location j.

People and cabs are matched via an online platform that sets prices and commission rates. People's willingness to pay is uniformly distributed in $[0, 1]$; so, if the platform sets price p_i at location i then there are $r_i = y_i(1 - p_i)$ riders willing to hire a cab at that location. The remaining people are assumed to use public transport or other means of travel, and they do not generate any revenue for the platform. The platform's software identifies cabs and passengers at location i and creates matches according to $m_i = \min\{r_i, x_i\}$. Cabs can accommodate only a single passenger per trip and the assignments are random; thus, the probability that a driver who is searching at location i finds a passenger is equal to

$$\eta_i = \frac{m_i}{x_i} = \min\left\{\frac{r_i}{x_i}, 1\right\}.$$

Occasionally, we refer to η_i as the utilization rate at location i, because from the platform's point of view η_i represents the percentage of cabs utilized in a ride.

[7]It is intuitive that fairness-sensitive customers would respond negatively to location-specific pricing, thereby undermining the performance of Model 4. A similar effect would likely arise in extensions of models such as Bimpikis et al. (2019), where origin- or origin-destination-based pricing can likewise deter fairness-minded users and reduce the attractiveness of such pricing schemes. However, the magnitude of this effect is not immediately clear, nor is it obvious whether it is substantial enough to warrant a change in the platform's policy choice. In our model, the performance gap between Model 4 and Model 3 (uniform price, flexible commission) is typically small, especially when the city layout and traffic flows are relatively uniform. Numerous simulations—based on both real-world ride patterns in New York and Los Angeles, and randomly generated city structures—confirm this finding. As a result, only a small share of fairness-sensitive customers is enough to overturn the platform's preference for Model 4. This is an actionable implication for platform design: it shows that the presence of such customers, even in small numbers, should not be ignored when choosing between flexible and uniform pricing regimes.

In addition to matching passengers to cabs, the platform sets prices and commission rates across the city. In terms of notation, p_i refers to the per-mile price associated with rides originating from location i. Similarly, the commission rate c_i refers to the percentage of the revenue that the driver takes home after completing a ride originating from location i. Drivers participate in this market if their expected earnings are greater than or equal to their outside option, w, which is the wage they could earn in the labor market. For now the parameters θ and w are exogenous. Later, in Section 5, we will relax this assumption by allowing free entry.

Each period starts with a matching session in which vacant cars at each location are matched with passengers. We ignore operating costs (petrol, insurance, etc.) as one can redefine the outside option net of such expenses. In line with recent studies on driver motivation and retention (Hall and Krueger, 2018; Allon et al., 2023b), we assume that drivers are more likely to be discouraged and quit if they struggle to find passengers. This can be in the form of, for instance, a driver simply switching off the app and logging on to another one, e.g., a driver switching from Uber to Lyft, commonly known as multi-homing in the literature (Allon et al., 2023a). In contrast, successful matches enhance their likelihood of continuing to offer services. More specifically, drivers matched with a passenger complete their journey, and continue to the next period with probability φ_h. Drivers without a match, on the other hand, continue with probability φ_l. We assume that $\varphi_h > \varphi_l$, i.e., being matched increases the likelihood of remaining in the service while being unmatched increases the likelihood of dropping out. At the end of each period, passengers reach their destinations, matches are dissolved, and the process starts again. When deciding where to search, drivers not only take into account the probability of finding a customer η_i, but also the price p_i, the commission rate c_i, and the average trip length originating from that location. Below we analyze their problem.

3.2 Drivers' Problem and the Steady State Equilibrium

Let V_i denote the value of searching at location i before the matching session starts. We have

$$V_i = \eta_i \sum_{j=1}^{n} a_{i,j}(c_i p_i \delta_{i,j} + \varphi_h V_j) + (1 - \eta_i)\varphi_l \max\{V_j\}_{j=1}^{n}.$$

With probability η_i, the driver is assigned to a passenger, and with probability $a_{i,j}$, the passenger travels to location j. If the driver agrees to take this trip, then his payoff is equal to the share of the revenue $c_i p_i \delta_{i,j}$ plus the value of search at location j. With probability $(1 - \eta_i)\varphi_l$, he gets no passenger and does not drop out, in which case, again, he obtains the value of search. A cab that is unable to get a passenger can move to another location if it is more advantageous to search there, which is why $\max\{V_j\}_{j=1}^{n}$ appears in the last expression. Drivers are allowed to refuse a match, but if they do so, they must wait until the next round, i.e., they cannot instantaneously re-enter the matching pool to draw a better ride. (Below we show that, in equilibrium, drivers do not turn down a match and go empty in search of a better opportunity.)

Steady State Equilibrium. We focus on a steady state in which the number of incoming rides to a location is equal to the number of outgoing rides from that location, i.e.

$$m_i = a_{1,i}m_1 + ... + a_{n,i}m_n, \text{ for all } i.$$

The left-hand side represents the outflow from i, whereas the right-hand side is the inflow into i. Since each ride consists of one passenger and one driver, the equation ensures that in the steady state, the number of passengers and the number of drivers at each location remain unchanged.[8] Some drivers may drop out, but since they are replaced one for one, this process does not affect the equation above. The relationship holds across the entire city, so letting $\mathbf{m} = (m_1, ..., m_n)$ we write

$$\mathbf{m} = \mathbf{m}T. \tag{1}$$

Lemma 1 *The number of rides in the steady state satisfies*

$$m_i = \sigma_i M, \tag{2}$$

where $M = \sum_{j=1}^{n} m_j$, *and* $\sigma > \mathbf{0}$ *is the unique steady state vector of the transition matrix* T.

All proofs are in the E-Companion. If there are a total of M moves in the city, then a fraction $\sigma_i \in (0,1)$ of those moves must be originating from location i. In the steady state, the incoming and outgoing traffic flows are equal to each other, so an alternative interpretation of (2) is that a fraction σ_i of the traffic must be directed towards i. Either way, the parameter σ_i is a proxy of how attractive/busy the location is. If σ_i is high, then we infer that location i is busy as it pulls in and sends out a significant amount of traffic.

Furthermore, we assume that in the steady state, drivers are indifferent across locations, i.e.

$$V_1 = ... = V_n = V. \tag{3}$$

The indifference condition implies that $\max\{V_j\}_{j=1}^{n} = V$; thus in equilibrium, drivers have no strict incentive to relocate and search at another location.[9] Our notion of the steady-state is characterized by $[i]$ the stationarity of the distribution of people across locations, captured by equation (1), and $[ii]$ drivers' indifference across locations, captured by equation (3). Temporary or even cyclical imbalances in the flow of traffic due to, say, rush hours, bad weather, football games, etc., may violate these conditions. Ignoring such fluctuations, we take a rather long-term view of the market and posit that the number of passengers at each location remains intact; thus $[i]$ must hold. Likewise, there should not be a lasting difference in expected profits across locations, affirming that condition $[ii]$ must be maintained. This notion of equilibrium is common in dynamic search and matching models, e.g. Lagos (2000), as it yields analytically tractable results.

Simplifying the expression for V_i, we have

$$V_i = V = \frac{\eta_i p_i c_i d_i}{1 - \varphi_l - \eta_i(\varphi_h - \varphi_l)} \text{ for all } i, \text{ where } d_i = \sum_{j=1}^{n} a_{i,j}\delta_{i,j}$$

is the average trip length of a ride originating from i. We label the locations 1 to n such that

$$d_1 < d_2 < ... < d_n,$$

i.e., location 1 has the shortest expected trip length, whereas location n has the longest. (We ignore

[8]People who have a low willingness to pay choose alternative methods of transportation. Though we remain agnostic about such passengers, we implicitly assume that the number of such passengers entering and exiting is the same, ensuring that equation (1) is sufficient for maintaining a steady state.

[9]In equilibrium, drivers do not turn down a match and go empty in search of a better opportunity. To see why, note that if a driver idles, he walks away with $\varphi_l V$, whereas if he accepts a match, then he obtains $p_i c_i \delta_{i,j} + \varphi_h V$. The second expression is larger than the first; thus, no driver idles voluntarily.

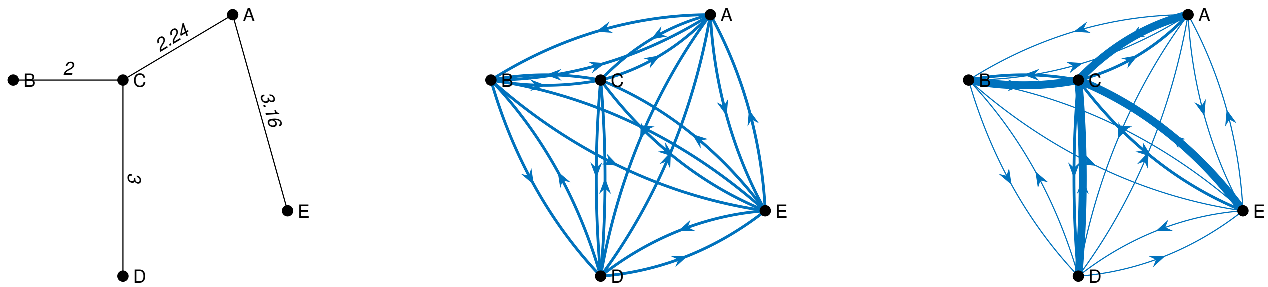

Figure 1: Layout and traffic flows

equalities.) Note that d_i depends on both on $\delta_{i,j}$ and $a_{i,j}$; thus, a location does not have to be physically the closest to other locations to have the smallest d_i.

Example. Consider the city in Figure 1 (left panel) and suppose that the transition probabilities are equal to each other (middle panel), i.e. $a_{i,j} = 1/4$ and $a_{i,i} = 0$. The expected trip lengths are equal to $d_A = 3.7$, $d_B = 4.7$, $d_C = 3.2$, $d_D = 5.4$, $d_E = 6.1$.[10] Location C sits at the intersection of routes and, as expected, it has the shortest trip length; thus, it can be labeled as location 1. A ranks second; so it is 2, B is 3, D is 4, and E is 5. Now suppose C attracts more traffic than other locations, e.g., suppose that 70% of traffic out of any location is directed towards C, while the remaining 30% is shared equally between the other three locations (i.e., $a_{i,C} = 0.7$ and $a_{i,j} = 0.10$, $i, j \neq C$). As for the traffic out of C, suppose it is still equally shared across the four destinations, i.e., $a_{C,i} = 0.25$. These flows are depicted in the right panel of Figure 1. The expected trip lengths are now equal to $d_A = 2.8$, $d_B = 3.1$, $d_C = 3.2$, $d_D = 4.0$, $d_E = 5.7$. Now A and B, despite being physically more remote, have shorter trip lengths than C. This is because when calculating d_A and d_B, shorter distances have a weight of 70% whereas longer distances have only 10%. The imbalance in the traffic flow changes the ranking; so, now location A ought to be labeled as 1, B as 2, C as 3, D as 4, and finally E as 5.

The re-labeling is important for the following reasons. From a driver's perspective the location with the minimum d_i is the least desirable location, and depending on the operating model, this location turns into a bottleneck if there are not sufficiently many cars in the city. The bottleneck location is typically the most central one–the one with a short physical distance to every other location–however, as the example illustrates, this is not always the case. Additionally, the location of the bottleneck can shift as the traffic flow changes. In the paper, we use numerical labels to identify locations, but it is important to note that these labels are relative and may vary if the transition matrix or city layout changes (for example, due to road closures or the creation of new roads that create new connections).

Drivers participate if $V \geq w$, i.e. if their expected earnings are at least as good as their outside option, w. The platform will not pay more than w, thus

$$V = w \Leftrightarrow \eta_i[p_i c_i d_i + (\varphi_h - \varphi_l)w] = w(1 - \varphi_l). \tag{4}$$

[10] Consider location A and note that $\delta_{A,B} = 4.24$, $\delta_{A,C} = 2.24$, $\delta_{A,D} = 5.24$, and $\delta_{A,E} = 3.16$. Since $a_{i,j} = 1/4$, the expected length of a trip originating from A is equal to $d_A = (4.24 + 2.24 + 5.24 + 3.16)/4 = 3.72$. Other trip lengths can be calculated similarly.

After substituting for η_i the equality becomes

$$m_i[p_i c_i d_i + (\varphi_h - \varphi_l)w] = x_i w(1 - \varphi_l), \text{ for all } i. \tag{5}$$

Combining (2) and (5) with the fact that $\sum_{i=1}^{n} x_i = \theta$ we obtain

$$x_i = \frac{\sigma_i p_i c_i d_i + \sigma_i(\varphi_h - \varphi_l)w}{\sum_{j=1}^{n} \sigma_j p_j c_j d_j + (\varphi_h - \varphi_l)w}\theta, \tag{6}$$

which pins down the number of cars at location i as a function of prices, commission rates, and expected trip lengths. Drivers prefer locations that are more busy (high σ_i) and that have longer trip lengths (high d_i). In addition, x_i rises in the price p_i and the commission rate c_i; thus, the platform can encourage drivers to search at location i by raising p_i or c_i. Such decisions are part of the platform's problem, which we study next.

Platform's Problem. The platform's per-period profit is equal to

$$\pi = \sum_{i=1}^{n} m_i p_i d_i - \sum_{i=1}^{n} m_i p_i c_i d_i.$$

The first term is the revenue generated through rides, the second term is the payout to drivers. Using the indifference condition in (5) the second term becomes

$$\sum_{i=1}^{n} m_i p_i c_i d_i = w(1 - \varphi_l)\theta - (\varphi_h - \varphi_l)wM. \tag{7}$$

Thus

$$\pi = \sum_{i=1}^{n} m_i p_i d_i - w\left[(1 - \varphi_l)\theta - (\varphi_h - \varphi_l)M\right] \tag{8}$$

The substitution of (7) eliminates commission rates from the platform's objective function. The platform picks prices to maximize the revenue while the commission vector $\mathbf{c}$ ensures drivers' participation and indifference via (5). Occasionally, we refer to such a $\mathbf{c}$ as *incentive compatible.* Finally, the platform's lifetime profit is equal to $\pi/(1 - \beta)$, where β is the discount factor.[11]

A steady state equilibrium is a time-invariant tuple $\{(p_i, c_i, x_i, y_i)\}_{i=1}^{n}$ such that (i) the platform maximizes its lifetime profit; (ii) drivers participate, and they are indifferent across locations; (iii) the inflow of moves equals to the outflow at each location; (iv) the total measures of cabs and passengers are equal to θ and 1, respectively.

For now, we treat the number of drivers θ as an exogenous parameter. Later in Section 5, we relax this assumption with free entry and determine the optimal fleet size. In what follows, we will analyze four different operating models:

- Model 1: Single Price, Single Commission Rate: $p_i = p$ and $c_i = c$ for all i.
- Model 2: Multiple Prices, Single Commission Rate: p_i is location specific, but $c_i = c$ for all i.
- Model 3: Single Price, Multiple Commission Rates: $p_i = p$ but c_i is location specific

[11] To save on notation, we leave out the discount factor in drivers' payoff calculations. The drop-out probabilities $1 - \varphi_h$ and $1 - \varphi_l$ serve as substitutes for discount factors, guaranteeing that drivers' lifetime payouts remain finite.

- Model 4: Multiple Prices, Multiple Commission Rates: Both p_i and c_i are location specific.

4 Operating Models

4.1 Model 1: Single Price, Single Commission Rate

We start with the claim that the probability of finding a customer, η_i, can be equal to 1 only at location 1, the location with the minimum distance. At remaining locations η_i must be strictly less than 1. To see why, note that since $c_i = c$ and $p_i = p$, the indifference condition (4) boils down to

$$pc(\eta_i d_i - \eta_j d_j) + (\eta_i - \eta_j)(\varphi_h - \varphi_l)w = 0 \text{ for all } i, j.$$

If $d_i < d_j$, then considering that $\varphi_h > \varphi_l$, the equality is possible only if $\eta_i > \eta_j$. Since $d_1 < ... < d_n$, only η_1 can be equal to 1, while all other η_is must be strictly less than 1.

We say there is *excess demand* at location i if $x_i < r_i$, i.e. if there are fewer cars than passengers. The previous claim rules out the possibility of excess demand at locations $i = 2, ..., n$. At location 1, however, η_1 may be equal to 1, which occurs when $x_1 = r_1$ or when $x_1 < r_1$. The following Lemma rules out the latter scenario.

Lemma 2 *There cannot be an equilibrium in which $x_1 < r_1$.*

In the proof, we start with a scenario with excess demand, and show that the platform is better off by increasing the price to absorb the excess demand and match it with the local supply. Before moving forward, let us introduce the following notation:

$$\mathbb{E}_\sigma(d) = \sum_{i=1}^{n} \sigma_i d_i \qquad \text{and} \qquad \mathbb{E}_\sigma(\sqrt{d}) = \sum_{i=1}^{n} \sigma_i \sqrt{d_i}.$$

The expression $\mathbb{E}_\sigma(d)$ is a weighted sum of d_is and can be interpreted as the average trip length in the city. $\mathbb{E}_\sigma(\sqrt{d})$ is similar.

Proposition 1 *If $\theta > \bar{\theta}_1$ then we have an interior equilibrium where the platform sets p^{int} and c^{int} and all locations exhibit excess supply. If $\theta \leq \bar{\theta}_1$ then the platform sets p^{cor} and c^{cor}.*[12] *In this corner equilibrium, the local supply matches the local demand at location 1, however, there is excess supply at the remaining locations.*

If there are sufficiently many cars in the city, i.e., if $\theta > \bar{\theta}_1$, then the platform optimally sets the interior price p^{int}. Along this outcome, all locations have more cars than passengers, but there is a caveat. Location 1, with its shortest expected trip length, sees fewer cars per passenger compared to any other location. Conversely, location n attracts more cars per passenger due to its longer trip length. The other locations fall somewhere in between these two extremes.

If the number of cars is insufficient, $\theta \leq \bar{\theta}_1$, then the platform experiences excess demand at location 1, so it resorts to a price increase to alleviate the excess demand (see Figure 2, left panel). In this corner equilibrium, the ratio of customers to cabs satisfies $\eta_i < \eta_1 = 1$, i.e. the local supply and demand are balanced at location 1—the least desirable location for drivers—but there is an oversupply of cabs at other locations. This outcome is wasteful from the platform's perspective. If

[12] Analytic expressions for $\bar{\theta}_1$, p^{int}, c^{int}, etc. can be found in the proof of Proposition 1 in the E-Companion.

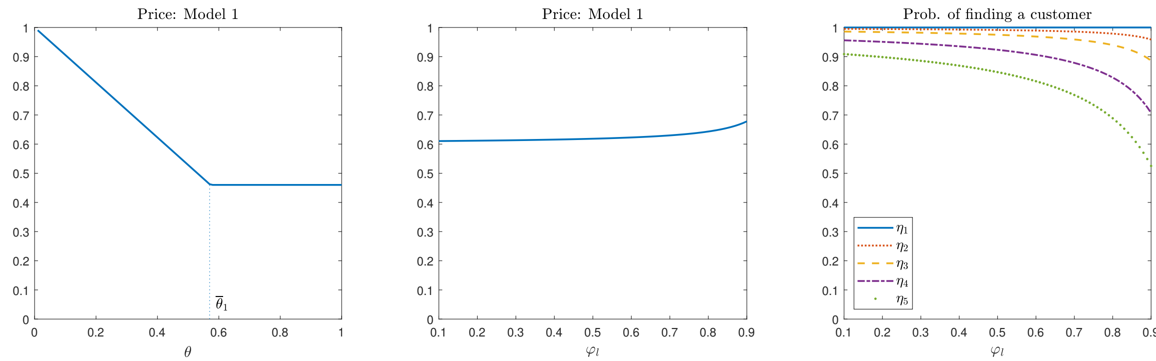

Figure 2: Model 1

the platform could somehow—e.g., by adopting a flexible commission model like Model 3—redirect surplus drivers to location 1, then it could avoid unnecessary price increases and cater to more customers. However, due to the inflexibility of the operating model, it cannot do so, resulting in location 1 becoming a bottleneck.

Driver Retention and Pricing Policy. Recall that in our model, drivers who successfully find customers are more likely to continue to drive, whereas drivers unable to find a match are more prone to dropping out. In what follows we explore how the platform reacts to this.

Remark 1 *A reduction in φ_l results in a decrease in price and an increase in the likelihood of finding a customer across all locations.*[13]

The Remark establishes that as φ_l decreases—indicating a heightened risk of drivers becoming discouraged and quitting—the platform lowers the price.[14] This adjustment is driven by the fact that lower prices stimulate more demand, improving the chances of drivers finding customers. This can also be seen in the middle and left panels of Figure 2: As φ_l decreases, the platform reduces the price (middle panel), which then increases the probability of finding a customer η_i across all locations (right panel). The higher η_is, in turn, help prevent drivers from dropping out. Indeed the platform's success in retaining drivers depends on its ability to ensure a steady stream of ride requests for drivers, providing them with consistent opportunities to earn income and reinforcing their decision to remain active on the platform.

Note that, in contrast to what one might expect, increasing prices does not lead to higher driver retention. Drivers drop out due to a lack of customers, and the platform addresses the root issue by lowering prices to stimulate demand. This dynamic underscores the interplay between driver retention and pricing strategy. Finally, the Remark also holds true for operating models 2, 3, and 4. However, as the underlying intuition remains unchanged, we will not reiterate it for the remaining models.

4.2 Model 2: Multiple Prices, Single Commission Rate

Lemma 3 *There cannot be an equilibrium in which $x_i < r_i$ at any location i.*

[13]The proof of the Remark involves showing that $dp^{int}/d\varphi_l$ and $dp^{cor}/d\varphi_l$ are both positive while $d\eta_i/d\varphi_l$ is negative (analytic expressions for the prices as well as η_i are in the proof of Proposition 1). We omit this step as it is relatively straightforward.

[14]A decrease in φ_l could be due to competition from other platforms that drivers can easily switch to. For example, multi-homing drivers may opt to switch from Uber to Lyft if they cannot find enough attractive opportunities.

This is similar to Lemma 2 and it has the same intuition. In the proof, we show that in case of excess demand, the platform is better off by increasing prices at locations where demand outstrips supply. This helps reduce local demand while incentivizing drivers to move toward those locations. Per the Lemma $r_i \leq x_i \Leftrightarrow m_i = r_i$ for all i. Since $m_i = \sigma_i M$ and $\sum_{i=1}^{n} y_i = 1$, we have $M = h(\mathbf{p})^{-1}$ and $m_i = r_i = \sigma_i h(\mathbf{p})^{-1}$, where

$$h(\mathbf{p}) = \sum_{i=1}^{n} \frac{\sigma_i}{1 - p_i} \quad \text{and} \quad g(\mathbf{p}) = \sum_{i=1}^{n} \sigma_i p_i d_i.$$

Substituting for m_i, we have

$$\pi(\mathbf{p}) = [g(\mathbf{p}) + w(\varphi_h - \varphi_l)]h(\mathbf{p})^{-1} - w(1 - \varphi_l)\theta. \tag{9}$$

The platform solves

$$\max_{\mathbf{p}} \ \pi(\mathbf{p}) \ \text{ s.t. } \ r_i \leq x_i \text{ for all } i,$$

where the constraints $r_i \leq x_i$ follow from Lemma 3.

Proposition 2 *If $\theta > \bar{\theta}_{2,0}$ then the platform sets p_i^{int} and c^{int}. In this interior equilibrium, the constraints are slack, so there is excess supply at all locations. Equilibrium prices satisfy $p_1^{int} < p_2^{int} < ... < p_n^{int}$.*

The interior equilibrium emerges if there are sufficiently many cars in the city ($\theta > \bar{\theta}_{2,0}$). With a surplus of cars, the platform can set prices without being concerned with attracting drivers to undesirable locations. In such an equilibrium, prices satisfy $p_i < p_{i+1}$, i.e. the platform sets higher prices at locations with longer trip lengths. This relationship can also be seen in the left panel of Figure 3, where prices in the interior region $\theta > \bar{\theta}_{2,0}$ satisfy $p_1 < ... < p_5$.[15] The platform faces a standard trade-off between extensive and intensive margin effects. On the extensive margin, it generates more matches by lowering prices, whereas on the intensive margin, it raises more money from each ride by increasing prices. The intensive margin effect is stronger at locations with longer trip lengths; thus the platform sets prices satisfying the above relationship.

If θ falls below $\bar{\theta}_{2,0}$ then the interior demand cannot be sustained with the available cars in the city and the constraints $r_i \leq x_i$ start to bind. In what follows we demonstrate that the constraints become activated in an orderly manner, first at location 1, then at location 2, and so on.

Lemma 4 *Let λ_k denote the Lagrange multiplier associated with the constraint $r_k \leq x_k$ for location k. If $\lambda_k = 0$ then $\lambda_{k+1} = 0$. Similarly if $\lambda_{k+1} > 0$ then $\lambda_k > 0$.*

The Lemma says that if the constraint is slack at location k, then it must be slack at longer-distance locations. Conversely, if it binds at location k, then it must also bind at shorter-distance locations. The implication is that the constraints bind in an orderly fashion, starting at the location with the shortest trip length (location 1), then at the location with the second shortest trip length (location 2), and so on. Letting $k = 1, .., n$, we refer to *regime*$-k$ as the outcome in which the first k

[15] In the simulations, we consider a city with five locations. The layout of the city and the transition matrix are as in Figure 1, right panel. Later, we calibrate the model for New York City and Los Angeles using real-world ride patterns from Uber. The findings from these two sets of simulations are highly comparable in nature.

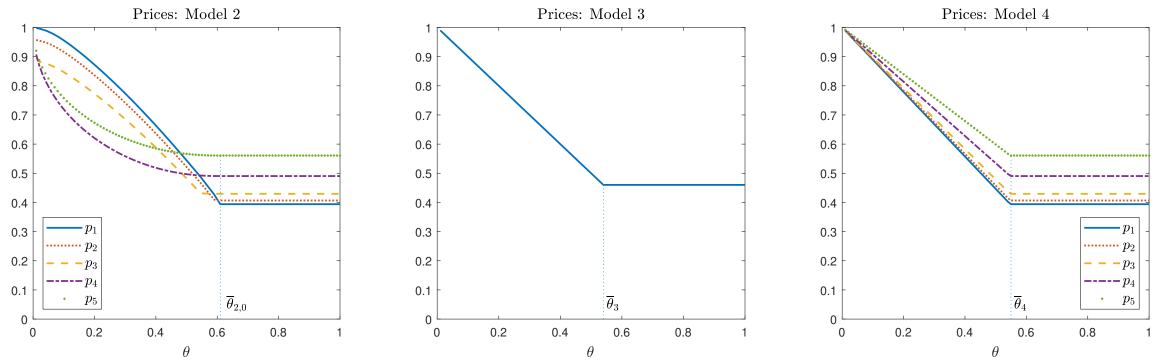

Figure 3: Prices in models 2, 3, 4

constraints are active. Prices in regime-k satisfy

$$p_{i,k}d_i[(1-\varphi_l)\theta h(\mathbf{p}_k)-(\varphi_h-\varphi_l)] = (1-\varphi_h)g(\mathbf{p}_k) \text{ for } i=1,...,k \quad \text{and} \tag{10}$$

$$[g(\mathbf{p}_k)+w(\varphi_h-\varphi_l)]\,h(\mathbf{p}_k)^{-1} = (1-p_{i,k})^2 d_i \text{ for } i=k+1,...,n. \tag{11}$$

The first set of equations follows from the binding constraints at locations 1 through k, while the second set arises from the first-order conditions at locations $k+1$ through n. The commission rate is then determined to satisfy the indifference condition (7).

Proposition 3 *Regime–k obtains if $\theta \in [\bar{\theta}_{2,k}, \bar{\theta}_{2,k-1})$. In this parameter region, there exists a set of feasible prices $p_{1,k},...,p_{n,k}$ and a commission rate c_k satisfying (10), (11) and (7). Prices satisfy $p_{1,k} > p_{2,k}... > p_{k,k}$ and $p_{n,k} > p_{n-1,k}... > p_{k+1,k}$, and are bounded below by $p_{\min} = 1+\theta/2-\sqrt{\theta^2/4+\theta}$. Finally, $p_{1,k}$ can be approximated by $p_{1,k} = 1-\kappa\theta$, where κ is given by (27), and the remaining prices can be pinned down via (25).*

The left panel in Figure 3 provides an illustration. If θ falls below $\bar{\theta}_{2,0}$ (about 0.6 in the simulation), then the platform runs out of cars at location 1, and the constraint $r_1 \leq x_1$ binds. In response, the platform increases p_1, which lowers the local demand r_1 and increases the local supply x_1 (by encouraging more drivers towards that location). The price intervention matches the local demand and supply, so $\eta_1 = 1$. At other locations, however, η_is are still less than 1. If θ falls further, e.g. below 0.57 in the simulation, then p_2 starts to rise in location 2, and if it falls below 0.52 then p_3 starts to rise in location 3 to match the local demand and supply. Even though the model cannot avoid bottlenecks, it still manages them locally by using location-specific prices, and as a result, it does not disturb the demand at other locations too much. For instance, when θ falls below $\bar{\theta}_{2,0}$, the price at location 1 surges up, but prices at remaining locations stay relatively unchanged.

The above observations seem to resonate with the surge pricing strategy employed by Uber. The surge pricing scheme kicks in when the number of passengers asking for a ride at a location exceeds the number of available drivers at that location, which in our model is equivalent to the constraint $r_i \leq x_i$ becoming active. Uber executives defend the surge pricing practice saying it serves their goal of "relentless reliability to manage the marketplace math so that supply and demand match as perfectly as possible in the face of ever-shifting, highly unpredictable circumstances" (Wohlsen, 2013). Our results seem to confirm a similar insight. Indeed when p_1 surges up at location 1, the local demand r_1 goes down while the local supply x_1 goes up. Furthermore, Uber's practice is location-specific, i.e., while the price may surge at an excess demand location, it remains unchanged at other locations. This outcome is also similar to what we observe in our simulations.

In highlighting our model's insights that may be relevant to real practice, such as surge pricing, we should note the following caveat. Our model is based on a steady-state setting; as such, Figure 3 depicts steady-state equilibrium prices associated with different values of θ. Uber's surge pricing practice, on the other hand, appears to be a temporary and transitional solution; thus, it may not be directly comparable to our steady-state results.

4.3 Model 3: Single Price, Multiple Commission Rates

Lemma 3 remains valid, i.e. there cannot be an equilibrium in which a location exhibits excess demand. Save for some minor differences (instead of prices, the platform uses commission rates to incentivize drivers towards excess demand locations) the proof remains the same, so we only provide a sketch here. Fix some p and let $\mathcal{S}_0$ denote the set of locations in which demand is greater than or equal to supply, with at least one location exhibiting excess demand, and $\mathcal{S}_1$ the set of locations with excess supply. Suppose that $\mathcal{S}_1$ is non-empty. If the platform leaves the price p as well as the commission rates in $\mathcal{S}_0$ intact, but reduces the commission rates in $\mathcal{S}_1$, then some drivers in $\mathcal{S}_1$ would flow towards locations in $\mathcal{S}_0$, creating more rides there. If the reduction is infinitesimally small, then despite losing drivers to $\mathcal{S}_0$, none of the locations in $\mathcal{S}_1$ would fall into excess demand, and the remaining drivers would still be able to serve the initial demand. Overall, the platform would not lose any profits in $\mathcal{S}_1$, yet it would create more rides and more profits in $\mathcal{S}_0$ rendering the intervention profitable. Now suppose $\mathcal{S}_1$ is empty, i.e. at every location we have $x_i \leq r_i$ with at least one inequality strict. Since all cars are being used in a ride, and there is still some excess demand, the platform can earn more by increasing p to the point where demand equals to supply at every location, rendering the conjectured outcome a non-equilibrium. In conclusion, so long as there is excess demand in the city, the platform has a profitable intervention; thus, there cannot be an equilibrium in which $x_i < r_i$ at any i.

Since $r_i \leq x_i$, we have $m_i = r_i = y_i(1-p)$. Recall that $m_i = \sigma_i M$ and $\sum_{i=1}^{n} y_i = 1$, so

$$m_i = r_i = \sigma_i(1-p) \quad \text{and} \quad M = (1-p)\,.$$

Substituting these into (8) yields

$$\pi = (1-p)\,p\mathbb{E}_\sigma\,(d) - w[(1-\varphi_l)\theta - (\varphi_h - \varphi_l)\,(1-p)].$$

The commission vector $\mathbf{c}$ must be incentive-compatible, i.e. it should satisfy (7), which, after substituting for m_i becomes

$$(1-p)\left[p\sum_{i=1}^{n}\sigma_i c_i d_i + (\varphi_h - \varphi_l)\,w\right] = w\,(1-\varphi_l)\,\theta. \tag{12}$$

The constraint $r_i \leq x_i$, after substituting for r_i and x_i, and using the equality above is equivalent to

$$r_i \leq x_i \Leftrightarrow w\,(1-\varphi_h) \leq pc_i d_i.$$

Recall that in Model 2, the constraints became active in an orderly fashion, starting at location 1 and then at location 2, and so on. Here, this is no longer the case.

Lemma 5 *Fix p. Suppose there exists an incentive compatible* $\mathbf{c}$ *under which* $r_i < x_i$ *for* $i \leq k$ *and* $r_i = x_i$ *for* $i > k$*. Then there exists another incentive-compatible* $\hat{\mathbf{c}}$ *under which* $r_i < x_i$ *for all* i*.*

The idea behind the proof is this. We can generate a new $\hat{\mathbf{c}}$ by marginally shaving off the rates of $\mathbf{c}$ at locations where the constraint is slack (but without rendering any of these constraints binding) and marginally increasing the rates at locations where the constraint is binding. Thus, by construction, all constraints become slack under $\hat{\mathbf{c}}$. The Lemma rules out the possibility that $r_i < x_i$ for some locations and $r_i = x_i$ at other locations. Either the constraints are slack at all locations or they bind at all locations. We can now characterize the equilibrium.

Proposition 4 *If* $\theta > \bar{\theta}_3$ *then all locations exhibit excess supply, i.e.* $\eta_i < 1$ *for all* i*. The platform sets* p^{int}*, but the commission rates are indeterminate. If, however,* $\theta \leq \bar{\theta}_3$*, then* $\eta_i = 1$ *for all* i*, i.e. no cab idles at any location, while the platform sets* p^{cor} *and* c_i^{cor}*.*

If θ is sufficiently large, then the platform can experiment with a wide range of commission schemes and still keep the customer-to-cab ratio η_i below 1 at all locations. If, however, θ falls below $\bar{\theta}_3$, the platform becomes unable to circumvent the constraints due to the overall shortage of drivers. Consequently, η_i reaches 1 across all locations, resulting in the platform operating at full capacity throughout the city. As the driver supply fails to meet interior demand, the price inevitably begins to increase (see the middle panel in Figure 3).

In contrast to the preceding two models, Model 3 gives the platform the ability to avoid bottlenecks. Thanks to the flexible commission structure, the platform does not resort to a price intervention until the passenger-to-cab ratio is equal to 100% at every location. Up to that point, by adjusting the commission rates—increasing them at less desirable locations, decreasing them at more desirable locations, or a combination—the platform manages to spread the cars evenly and serve the demand associated with the interior solution. Thus, in contrast to the previous model, no location turns into a bottleneck.

4.4 Model 4: Multiple Prices, Multiple Commission Rates

Finally, we turn to the most flexible operating model. As before, we start with the claim that there cannot be an equilibrium in which $x_i < r_i$ at any i. (The proof is omitted as it closely resembles the previous ones.) Since $r_i \leq x_i$ we can write $m_i = r_i = y_i(1 - p_i)$ and since $m_i = \sigma_i M$, and $\sum_{i=1}^{n} y_i = 1$, we have

$$m_i = r_i = \sigma_i / h(\mathbf{p}).$$

Substituting for m_i, the platform's profit is equal to

$$\pi(\mathbf{p}) = [g(\mathbf{p}) + w(\varphi_h - \varphi_l)]h(\mathbf{p})^{-1} - w(1 - \varphi_l)\theta. \tag{13}$$

The commission vector $\mathbf{c}$ must be incentive-compatible in that it should satisfy drivers' indifference via equation (5), which, after substituting for m_i and x_i, becomes

$$\left[\sum_{i=1}^{n} \sigma_i p_i c_i d_i + (\varphi_h - \varphi_l)\, w\right] h\,(\mathbf{p})^{-1} = w\,(1 - \varphi_l)\,\theta. \tag{14}$$

The constraint $r_i \leq x_i$, after substituting for r_i and x_i, and using the equality above is given by

$$r_i \leq x_i \Leftrightarrow w(1 - \varphi_h) \leq p_i c_i d_i \text{ for all } i.$$

Since the commission rates are flexible, either the constraints are slack at all locations or they bind at all locations. In other words, there cannot be a scenario where $p_i c_i d_i > w$ for some locations and $p_i c_i d_i = w$ at other locations. When faced with such an outcome, the platform can simply shave off the commission rates at excess supply locations and increase the rates at constrained locations to slacken those constraints. The proof of this claim is practically the same as the proof of Lemma 5; thus it is skipped here.

Proposition 5 *If $\theta > \bar{\theta}_4$ then all locations exhibit excess supply. The platform sets p_i^{int} but the interior commission rates are indeterminate. If, however, $\theta \leq \bar{\theta}_4$, then no location exhibits excess supply. Along this outcome the platform sets p_i^{cor} and c_i^{cor}.*

With sufficiently many cars in the city ($\theta > \bar{\theta}_4$), the equilibrium is interior and no constraint $r_i \leq x_i$ is active. Thanks to the flexible nature of the commissions, the platform avoids the constraints until $\theta = \bar{\theta}_4$. Up to that point, by fine-tuning the location-specific rates, the platform incentivizes the drivers to spread themselves across the city in an even way, and thereby, it avoids bottlenecks. If, however, θ falls below $\bar{\theta}_4$, then the interior demand cannot be addressed with the number of available cars, so prices start to rise (see the right panel in Figure 3).

Equilibrium prices satisfy $p_i < p_{i+1}$, i.e. the platform sets higher prices at locations with higher d_i. This relationship is similar to what we saw in Model 2 and shares the same underlying logic. In Model 2, the relationship broke down in the corner region because, in that model, prices at bottleneck locations had to surge up to incentivize drivers. Here, the commission rates are employed for this purpose; thus, the relationship remains valid both in the interior and the corner equilibria.

5 Free Entry

So far, we treated θ and w as exogenous parameters. We now relax this restriction by assuming that drivers are free to enter, and the labor supply is governed by

$$\theta = \mu w \tag{15}$$

where w, with some abuse of notation, is expected earnings and μ is a labor market sensitivity parameter, i.e., the higher the value of μ, the more responsive the supply is to a change in expected earnings. The equation indicates a positive relationship between earnings and entry meaning that higher prices or commissions attract more drivers to the platform. This positive relationship is backed by empirical research on labor supply in the gig economy. For instance, a recent study by Castillo (2023) examined Uber drivers' labor choices in the greater Houston area, discovering a positive relationship between earning opportunities and driving hours, both in the short-term and long-term.

Here we rely on the parameter μ to capture labor market responsiveness to earning opportunities. As a robustness check, in Section A.2 we present a more granular approach, where we distinguish between different groups within the potential driver pool, each varying in population size and responsiveness. This framework not only helps the platform determine optimal entry levels and driver compensation but also provides insight into the composition of total entry—specifically, how much each subgroup contributes—and how different groups adjust to changes in commission rates and prices.

In what follows, we analyze the selection of the optimal number of drivers θ_i^* under model i. Start with Model 4 and recall that the platform's profit function is

$$\pi(\mathbf{p}) = [g(\mathbf{p}) + w(\varphi_h - \varphi_l)]h(\mathbf{p})^{-1} - w(1-\varphi_l)\theta.$$

If $\theta > \overline{\theta}_4$ then the price vector $\mathbf{p}$ is given by (31) and if $\theta \leq \overline{\theta}_4$ then it is given by (32). In the proof of Proposition 6 we show that the relevant region to consider is $\theta \leq \overline{\theta}_4$[16], thus the platform solves

$$\max_{\theta} \pi(\mathbf{p}) \quad \text{s.t.} \quad (15) \text{ and } (32),$$

to obtain the optimal entry θ_4^*. We use the same procedure for the other models as well.[17]

Proposition 6 *Optimal entries are given by*

$$\theta_1^* = \frac{1}{2}\frac{\mu d_1 \left[\mathbb{E}_\sigma(d)(1-\varphi_h) + d_1(\varphi_h - \varphi_l)\right]}{\mu d_1^2(1-\varphi_l) + \mathbb{E}_\sigma(d)(1-\varphi_h)^2 + d_1(1-\varphi_h)(\varphi_h-\varphi_l)}, \tag{16}$$

$$\theta_3^* = \frac{\mathbb{E}_\sigma(d)}{2(1-\varphi_h)/\mu + 2\mathbb{E}_\sigma(d)} \quad \textit{and} \quad \theta_4^* = \frac{\mathbb{E}_\sigma(d)}{2(1-\varphi_h)/\mu + 2\mathbb{E}_\sigma^2(\sqrt{d})}.$$

The optimal entry under Model 2 is approximately equal to

$$\theta_2^* = \frac{d_1}{2(1-\varphi_l)/\mu + 2\kappa d_1}. \tag{17}$$

Fixed commission models generally lead to more entries than flexible commission models.

The platform picks prices and commission rates, while the number of entrants is pinned down through the indifference condition (14). A generous combination of $\mathbf{c}$ and $\mathbf{p}$ rewards drivers with high earnings, leading to a large number of drivers entering the market, while a less generous combination leads to the opposite. The equilibrium values of $\mathbf{c}$ and $\mathbf{p}$ ensure that the right number of drivers enter the market.

Fixed commission models, 1 and 2, generally lead to more entries than flexible commission models, 3 and 4. This is because flexible commission models can efficiently utilize all available vehicles in rides, leading to fewer cars being needed. Fixed commission models, on the other hand, create bottlenecks and under-utilize drivers in rides. Therefore, they require more cars to function, which means higher entry numbers. Higher entries, however, do not translate to increased matches or profits. We will revisit this point in Proposition 8.

Labor Market Sensitivity. A key factor that affects the number of drivers entering the market, as well as other equilibrium objects—prices, commissions, number of matches, etc.—is the labor market sensitivity parameter, μ.

Proposition 7 *A fall in μ results in higher prices and commissions, and in higher driver compensation overall. Additionally, it leads to lower entries, fewer matches, lower profits, and lower consumer surplus.*

[16] The intuition is this. In the region $\theta > \overline{\theta}_4$, there is an excess supply of drivers, leading to the under-utilization of some drivers, which is suboptimal. Therefore, the optimal entry must be in the region $\theta \leq \overline{\theta}_4$.

[17] We solve a sequential optimization problem where the parameter θ is initially fixed—as it was up to this point—and its optimal value is obtained subsequently.

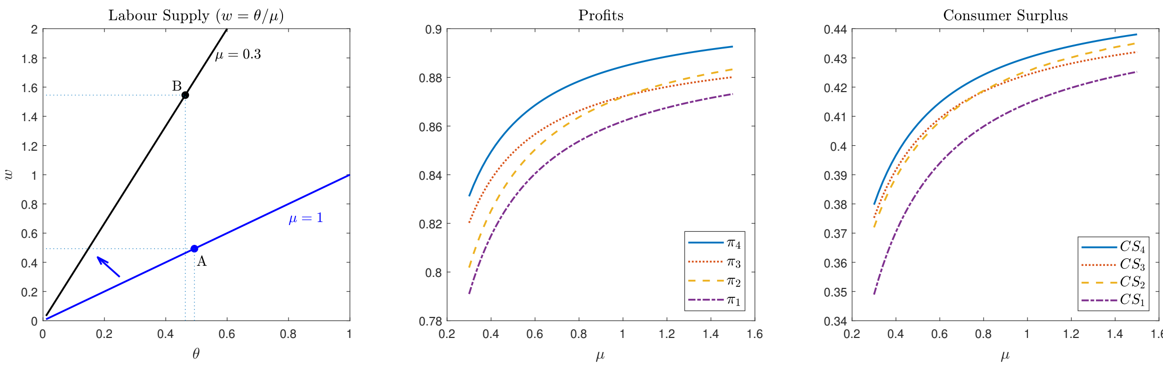

Figure 4: Labor Supply, Profits and Consumer Surplus

A decline in μ causes a rotation in the labor supply curve, indicating that fewer drivers would join the platform when offered the same level of compensation. To attract drivers, the platform responds by increasing prices and commission rates. However, despite the higher compensation, overall entry declines. Figure 4 (left panel) illustrates this effect. The labor supply curve is given by $w = \theta/\mu$. When $\mu = 1$, labor supply is relatively responsive, with the optimal entry, assuming Model 4, calculated as $\theta = 0.5$ and the corresponding expected compensation as $w = 0.5$ (point A in the figure). As sensitivity drops to $\mu = 0.3$, the labor supply curve rotates upward, becoming less responsive to earning opportunities. In response, the platform raises prices and commission rates significantly (not shown here) to attract drivers, resulting in a more than threefold increase in overall compensation, reaching $w = 1.54$ (point B). Despite this increase, the new entry θ falls slightly below the previous level. The remainder of the Proposition is straightforward to understand. Due to the higher prices, demand for rides falls compared to previous levels. As a result, fewer matches occur, leading to lower profits and smaller consumer surplus. These claims are also evident in the middle and right panels of Figure 4, showing that across all operational models, both profits and consumer surplus fall as μ falls.

Remark 2 *In response to changes in μ, the platform relies more heavily on commission adjustments than price adjustments. For instance, a decrease in μ results in a larger relative increase in commissions.*

In the proof we show that $|\epsilon(c_i, \mu)| > |\epsilon(p_i, \mu)|$, where $\epsilon(c_i, \mu)$ is the elasticity of the commission rates with respect to μ, and $\epsilon(p_i, \mu)$ is the elasticity of prices with respect to μ. Both elasticities are negative, hence, the inequality is expressed in terms of absolute values. The inequality establishes that the platform's adjustments to commissions are (significantly) more pronounced than its adjustments to prices.

For instance, if μ falls then potential drivers become less likely to join. The platform responds to this by offering a higher overall compensation level to maintain sufficient driver numbers. This is achieved by raising both commissions and prices, but the commission increase is significantly larger. The rationale for this is that raising prices suppresses customer demand, while raising commissions incentivizes drivers without unnecessarily affecting demand. A small price increase is still needed to eventually balance the demand with the reduced driver supply, but it is (much) smaller than the commission adjustment.

We explore this insight further in Section A.2 with a different model of labor supply and driver entry. Our simulation looks at how commission rates and prices change as the minimum wage drivers require to join the platform increases. In our current setup, this is equivalent to μ falling. The simulations (right panel of Figure 9) show that the platform responds to this with a modest price increase and a much larger increase in commissions.

A more general implication of the aforementioned results is that the platform must prioritize consideration of the labor market sensitivity specific to the cities or localities in which it operates. It should tailor its operating models, pricing structures and commissions based on these local factors, rather than relying on overarching, one-size-fits-all policies. Indeed, Uber already seems to be taking this approach: in its recent efforts to grow its driver base in London and across the UK to meet the growing demand, Uber has announced changes to its pricing and driver compensation policy which critically "vary city by city" (Uber, 2022b).

5.1 Profits, Consumer Surplus, and Social Welfare

The consumer surplus at location i is given by

$$cs_i = \sum_{j=1}^{n} y_i a_{i,j} \delta_{i,j} \int_{p_i}^{1} (v - p_i) dF(v)$$

where v is the willingness to pay, p_i is the price and $F(v)$ is the CDF governing v. Noting that $\sum_{j=1}^{n} a_{i,j}\delta_{i,j} = d_i$ and $F(v) = v$ (uniform distribution) we have

$$cs_i = \frac{1}{2} m_i d_i (1 - p_i).$$

The equation follows from the fact that, in equilibrium, we have $m_i = y_i(1 - p_i)$. The consumer surplus for the entire city is equal to

$$CS = \frac{1}{2} \sum_{i=1}^{n} m_i d_i \left(1 - p_i\right).$$

We can now compare the operating models based on the number of matches they generate, as well as the amount of profits and consumer surplus they produce.

Proposition 8 *We have (i) $M_4 > M_3 > M_1$, (ii) $\pi_4 > \pi_3 > \pi_1$, and generally (iii) $CS_4 > CS_3 > CS_1$. In other words, Model 4 outperforms Model 3 in terms of creating more matches, profits, and consumer surplus, and Model 3, in turn, outperforms Model 1. The performance of Model 2 is comparable to that of Model 3. When μ is small, Model 3 outperforms Model 2 by generating more matches, higher profits, and greater consumer surplus. Conversely, when μ is large, Model 2 outperforms Model 3.*

Flexible commission models, 3 and 4, make efficient use of all available vehicles, leading to higher performance. In contrast, fixed commission models, 1 and 2, suffer from inefficiencies such as bottlenecks and under-utilized drivers, resulting in fewer matches, lower profits, and reduced consumer surplus. The middle and right panels of Figure 4 provide an illustration for these insights. The Proposition establishes them analytically.

Model 4 stands out as the most versatile operating model and generally delivers the best performance across all metrics. On the other hand, Model 1 is the least flexible and tends to perform the poorest. Models 2 and 3 fall in between. While this observation may not be surprising, it is worth noting that in situations involving behavioral customers who may perceive location-specific pricing as "unfair", Model 3 might actually outperform Model 4. Below we will come back to this point.

One might wonder how, say, Model 4 can create more profits *and* more consumer surplus at the same time, as higher profits usually mean lower consumer surplus. This is possible because, due to its flexible nature, it creates more matches than other models, which translates into more profits and more consumer surplus. Said differently, it creates a bigger pie; thus higher profits do not necessarily equate to lower consumer surplus.

Driver Retention. Before moving to the next section, we revisit how changes in φ_l affect platform outcomes under free entry. Recall that a lower φ_l raises the likelihood that unmatched drivers become discouraged and leave the platform. In Models 1 and 2, this dropout risk prompts the platform to lower prices to stimulate demand and increase the matching rate. While this response reduces premature driver exit, it also lowers platform profits. In contrast, in flexible commission models, 3 and 4, the equilibrium objects—amount of entry, prices, commission rates, profits, matches—do not depend on φ_l at all.[18] This is because the flexible commission policy ensures full utilization of the driver fleet, so no driver remains unmatched in equilibrium. As a result, the risk of premature dropout has no bearing on the platform's pricing or profit.

The implication is that if the platform adopts a flexible commission policy, it can significantly reduce the risk of drivers leaving the market too soon. The flexible commission rule ensures that drivers are fully utilized, and as a result, the platform benefits from not having to deal with the disruption and challenges associated with the frequent turnover of drivers. This represents a significant operational advantage for the platform.

Social Welfare. We define social welfare as the sum of platform profits, consumer surplus, and driver surplus, i.e. $SW = \pi + CS + DS$. Driver surplus arises because the equilibrium wage is pinned down by the marginal driver—the one just indifferent between entering or not. Other drivers are willing to participate at lower earnings, so when they receive the equilibrium wage, they are effectively paid more than the minimum amount that would have induced them to join. This gap between what they actually earn and the level at which they would have participated represents an extra rent. Summing these rents across all such drivers yields the total driver surplus.

In this sense, driver surplus reflects the extra benefit received by drivers who would have been willing to participate even at lower earnings. Its magnitude depends on how many drivers enter and how the wage compares to their underlying participation thresholds. Since these elements interact in nontrivial ways with operating models, we evaluate driver surplus as part of the numerical simulations in the next section, when we analyze New York City and Los Angeles.

5.2 Behavioral Customers

Model 4 delivers the highest profits and is therefore, in principle, the natural choice for the platform. However, it relies on location-specific pricing, which results in higher prices in certain areas. This practice may be perceived as unfair by some passengers, potentially leading them to disengage from

[18]In Model 1, a decline in φ_l leads the platform to lower the price (see equation 21), which in turn reduces its profit. The same relationship holds in Model 2, where a lower φ_l reduces platform profit, given by (41). In contrast, in Models 3 and 4, driver entry (16), prices (30, 32), profits (38), and the remaining equilibrium objects are independent of φ_l.

the platform. As a result, Model 4's performance may fall short of uniform-pricing alternatives such as Model 3, particularly in markets where fairness concerns strongly influence consumer behavior.

Indeed, customers seem to develop mental *reference points* based on their past experiences or their expectations of what should be a "fair" price (Bolton et al., 2003). In the context of ride-hailing, these reference points may come from previous trips, competitors' prices, or the platform's fares in the past. When Model 4 creates relatively higher fares at certain locations, customers may perceive these deviations as losses, leading to dissatisfaction and calling such a practice as unfair. This sense of unfairness can be further compounded by *inequity aversion* (Fehr and Schmidt, 1999), where customers compare their fare to what others might be paying. Those who pay higher fares may think that riders in other locations are receiving better deals for basically the same service. This feeling of inequity may provoke emotional responses, including disengaging from the platform and seeking alternative options.

The response to Uber's surge pricing appears to support these ideas.[19] Despite its economic rationale, surge pricing has faced significant backlash, illustrating how price differences—whether due to short-term fluctuations or long-term structural factors—can lead to strong consumer dissatisfaction. It is described as "price gouging" by The New York Times (Lowrey, 2014) and "exploitative" by Harvard Business Review (Dholakia, 2015). CNN Business went further, labeling such pricing practices as feeling like a "scam"(Morrow, 2024). These negative perceptions have even led competitors to explore alternative strategies. For instance, Lyft introduced a \$2.99 monthly subscription service called Price Lock, which fixes fares on specific routes at select times. According to the company's CEO, this feature was designed to address what he called the app's "most hated feature"(CBS News, 2024). These concerns highlight a critical trade-off for ride-hailing platforms. If fairness concerns lead to a significant drop in consumer participation, the platform, then, needs to reconsider its pricing strategy.

Considering the presence of such customers, a uniform pricing approach, as in Model 3, could prove more effective. To explore this point in more detail, suppose that a proportion $\alpha \in (0, 1)$ of customers turn off their app when they realize that the app is using location-specific pricing.[20] In Model 4, due to the shrinking customer base, the new optimal entry is approximately equal to

$$\theta'_4 = (1 - \alpha)\theta^*_4.$$

To see why, note that the parameter θ represents the ratio of drivers to customers. If the measure of customers drops from 1 to $1 - \alpha$, then the amount of entry adjusts proportionally. With the reduced entry, the profit now shrinks to

$$\pi'_4 = (1 - \alpha^2)\pi_4,$$

which is obtained by substituting θ'_4 into the profit function in Model 4. (Expressions for π_3 and π_4 are given by (38) in the E-Companion.)

If, on the other hand, the platform were to use Model 3, then it would still earn π_3 because it is based on uniform pricing and, therefore, the parameter α has no impact. It is clear that if α is

[19] Surge pricing, as practiced by Uber, is based on real-time demand fluctuations, whereas Model 4's location-specific pricing in our model reflects long-term structural factors. Both strategies, however, result in price differences across locations, which is the primary driver of consumer dissatisfaction.

[20] To maintain the steady state, we implicitly assume that the proportion of such customers is equal to α across all locations, i.e., no location exhibits a significantly higher or lower proportion of such customers compared to others.

sufficiently large, then Model 4 under-performs Model 3. In particular one can show that $\pi_3 > \pi_4'$ if

$$\alpha > \sqrt{\frac{\mu[\mathbb{E}_\sigma(d) - \mathbb{E}_\sigma^2(\sqrt{d})]}{1 - \varphi_h + \mu\mathbb{E}_\sigma(d)}}.$$

The threshold on the right-hand side depends, among other things, on the traffic and the shape of a city. In cities where the traffic is balanced and locations have similar expected travel distances, $\mathbb{E}_\sigma(d)$ and $\mathbb{E}_\sigma^2(\sqrt{d})$ will be close to each other, so even a small α will cause Model 4 to under-perform Model 3. For instance, using the results from our calibration in the next section, the implied threshold for New York is approximately 10%, i.e., if more than 10% of customers are sensitive to fairness concerns, then the platform would be better off implementing a uniform pricing policy rather than a flexible one. The implication is that in cities where fairness sensitivity is prevalent, adopting a uniform pricing structure could improve customer engagement and long-term profitability.

A natural question is how one might estimate α, the share of fairness-sensitive consumers. One can consider two complementary approaches. First, discrete choice surveys or experiments could present respondents with ride options with and without location-specific pricing, allowing researchers to quantify the share of fairness-sensitive consumers. This approach has precedent in the behavioral operations and transportation literature, including studies of fairness perceptions in car ownership and road pricing schemes (Schuitema et al., 2011). Second, one could use rich transaction-level data from a platform that has experimented with different pricing policies across locations. By comparing observed rider choices before and after the introduction of such pricing—while controlling for supply-side and demand-side factors—it would be possible to back out the fraction of users whose behavior is consistent with fairness sensitivity.

In the preceding analysis, we focused on comparing Models 3 and 4 while intentionally excluding 1 and 2. To see why, note that Model 1 uses uniform pricing, which avoids the issue of behavioral customers. However, Model 3 also has uniform pricing, avoids the behavioral customer issue, and, as shown in Proposition 8, outperforms Model 1. Therefore, further consideration of Model 1 is unnecessary. Model 2, on the other hand, relies on location-specific pricing and thus faces the same challenges with behavioral customers. Since Model 4 encompasses Model 2 as a special case and performs better, Model 2 does not warrant further consideration either. For these reasons, when accounting for behavioral customers, the relevant comparison is the one between Models 3 and 4.[21]

A Brief Discussion on Fairness Concerns. The literature on fairness in behavioral economics is vast, and we do not attempt a comprehensive review in our study. Instead, we highlight several key contributions that have shaped the way fairness considerations are understood in pricing and market behavior, and we take these insights as a guide in analyzing our platform pricing setting.

Fairness concerns often shape behavior in market transactions and in many cases act as a "behavioral constraint" for firms. Kahneman et al. (1986) show that consumers evaluate price changes not only by their economic impact but also by their perceived fairness. Raising a price is

[21] Even though Models 1 and 2 are ultimately dominated in terms of outcomes, they are still worth examining. These models serve as important benchmarks that help clarify the role of flexibility in platform design. By starting from the least flexible setting and gradually introducing location-specific pricing and commissions, we can identify the issues that arise when the platform lacks flexibility—such as the emergence of bottlenecks, under-utilization of the driver fleet, or the need for demand-distorting price interventions. By comparing these outcomes with those under Models 3 and 4, we can identify the operational value of introducing flexibility in pricing or commissions. We thank an anonymous reviewer for prompting us to explore this point.

deemed acceptable if it maintains a certain amount of profit (for example, to cover higher costs), but the same price increase is widely judged unfair if it is seen as exploiting a surge in demand. Such perceptions effectively impose a constraint on sellers—hence the title of their seminal paper.

When pricing decisions overlook this behavioral constraint, the outcome can be costly: customers may abandon the relationship, turn to competitors, spread negative word-of-mouth, or take other actions that undermine the firm. Xia et al. (2004) document these effects and emphasize that managers must be mindful of customer reactions when setting prices. Survey evidence confirms that practitioners themselves recognize this constraint. For example, Eyster et al. (2021) summarize responses from more than 12,000 firm managers and find that considerations of "implicit contracts" with customers consistently rank among the most important explanations for price rigidity. Managers frequently report implicitly stabilizing prices "out of fairness to customers," with such concerns receiving the highest median rank among competing theories.

Taken together, these studies establish that fairness concerns act as a fundamental behavioral constraint on pricing policy. Our paper builds directly on this principle and applies it to the context of ridesharing and platform policy design. Indeed, our analysis shows that fairness concerns influence the effectiveness of alternative platform policies, highlighting the interaction between operational decisions and perceived fairness.

5.3 Calibration

In what follows, we calibrate the model for New York City and Los Angeles based on real-world ride patterns. Before giving the details of our calibration, we provide a brief discussion on the background of ride-hailing in both cities. Uber began operating in New York City in 2011, launching in Manhattan as one of its first major expansions beyond San Francisco. Lyft entered the market in 2014, initially serving Brooklyn and Queens (Luckerson, 2014). However, in Manhattan—particularly Midtown, the Upper East Side, Chelsea, and the West Village—yellow cabs remain competitive due to high street-hailing demand. Indeed, these neighborhoods account for a disproportionate share of total ride activity in both the taxi and ride-sharing datasets which we discuss below.

Uber launched in Los Angeles in 2011, with Lyft following in 2013. Unlike New York, LA had a weaker incumbent taxi industry and little tradition of street-hailing, allowing ride-sharing to scale rapidly. The city's dispersed layout—with major hubs such as Downtown LA, West Hollywood, Santa Monica, and LAX separated by longer travel distances—leads to a more imbalanced traffic structure than in NYC. This difference has implications for platform performance in our model which we explore in more detail below.

Our data is extracted from a publicly available connectome map of rides on Uber's website. (Uber, 2019; Bimpikis et al., 2019). Much like a classical connectome map showing point-to-point spatial connectivity of neural pathways in the brain, the connectome map that was available on Uber's website included a visual map of the ride patterns during July 2014 among the neighborhoods of these cities. Actual ride frequencies, however, were not readily available in this visual map. Using the open html code of the website, we were able to scrape the data to obtain raw details such as borders defining various neighborhoods in both cities (much like the n locations in our model), the name of each neighborhood, the latitude/longitude coordinates defining the center of each neighborhood, as well as the weights of links between each coordinate, proxying the relative likelihood of a ride going from one coordinate to the other. We then created the transition matrix T

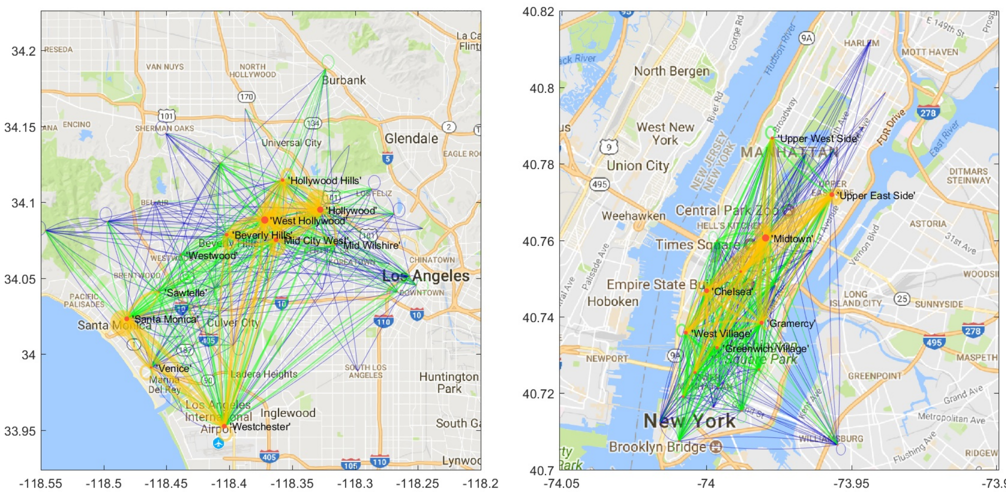

Figure 5: Traffic Flow Maps in NYC and LA

using the ride patterns between the nodes, and the distance matrix via Google Maps API using the coordinates of the nodes (see the E-Companion for the transition and distance matrices for NYC).[22]

To visualize the traffic flows, we constructed network maps using the scraped data, which include neighborhood-level ride flows and the geographic coordinates of each location. We interpret the transition matrix T as a weighted directed graph, where each edge $a_{i,j}$ reflects the traffic intensity from location i to j. Line thicknesses are proportional to these weights, while node sizes reflect the steady-state mass σ_i, capturing how busy each location is. The resulting network is overlaid on actual city maps using the corresponding geo-coordinates.

Figure 5 presents the resulting visualizations. In both cities, blue lines indicate light traffic, green lines moderate flows, and yellow lines heavy traffic. Larger orange dots indicate busier locations. In New York, the busiest areas include Midtown, Upper East Side, Chelsea, and West Village. In Los Angeles, high-traffic zones include West Hollywood, Beverly Hills, Downtown LA, Santa Monica, and Westchester (LAX). While New York displays a relatively balanced and interconnected network, Los Angeles exhibits more polarized flows, with dense west-side activity and sparse connections elsewhere. These structural differences have important implications for the relative performance of alternative operating models.[23]

[22]For NYC there were 29 locations, so our data-scraping process yielded a 29-by-29 matrix containing the weights of links across all locations, proxying the strength of the flow of traffic to and from each location. Out of 29 locations, there were 4 locations with a zero row or column, which meant that during that relatively short period, there was no traffic to or from those locations. We deleted those locations, so the final weight matrix was based on 25 locations. We recovered the transition matrix T from this weight matrix by normalizing the sum of the row vectors to 1. In addition, in both cities, some diagonal elements of the transition matrices were non-zero, which means that several rides that started and ended within the same neighborhood, e.g. Upper East Side in NYC. Similarly some non-diagonal elements were zero, indicating no rides took place between those locations. These facts violate our assumptions that $a_{i,j} > 0$ and $a_{i,i} = 0$, but they do not affect the inner workings of the model. Our assumptions are sufficient, but not necessary, to ensure that the transition matrix has a unique steady state vector σ. Our calculations show that the transition matrices associated with both cities are ergodic; thus they both have unique steady-state vectors. Finally, the distance matrix in our model is symmetric, i.e. $\delta_{i,j} = \delta_{j,i}$, however on occasion the distances between points i and j returned by the Google Maps API varied depending on which one was chosen as the origin and destination. We ignored such minor differences and used the shortest distance instead.

[23]To assess the robustness of our traffic patterns, we also analyzed the official yellow cab data from New York City for March 2024, made available by the NYC Taxi & Limousine Commission. The dataset contains over 3.5 million trips. After matching pickup and dropoff zone codes to location names, we constructed a full trip matrix across zones.

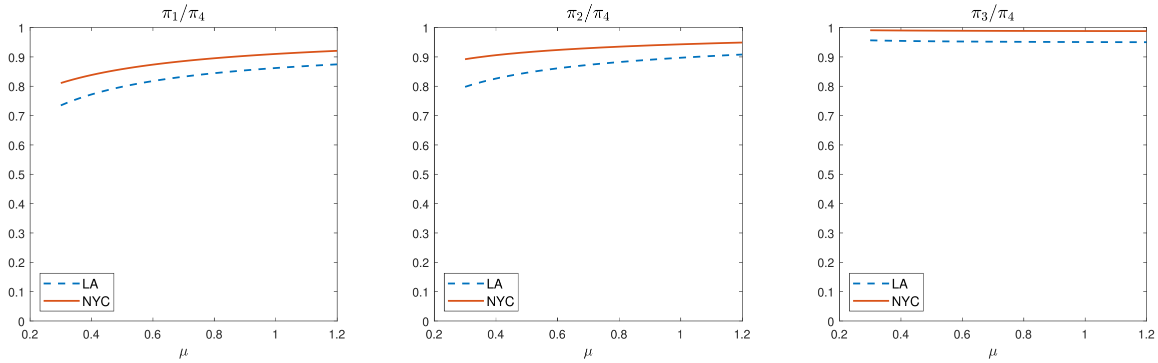

Figure 6: Performance of operating models in NYC and LA

Based on the transition matrix T we next calculate the platform's profit under each model and compare their performances. Figure 6 depicts equilibrium profits against the labor market sensitivity μ. Model 4 is the most versatile operating system and encompasses the other models as special cases; as such π_4 serves as a benchmark in the simulations. The panels illustrate π_1, π_2 and π_3 as a fraction of π_4. Note that we assume no behavioral customers in the simulations.

The left and the middle panel in Figure 6 show the performance of fixed commission models (1 and 2). In NYC, they yield profits 10% to 20% lower than the benchmark, with even lower outcomes observed in LA. The right panel shows the performance of Model 3, and interestingly, the under-performance is not significant, with a difference of less than 2% in NYC and less than 5% in LA. This highlights the importance of flexible commissions in avoiding bottlenecks and distributing the drivers evenly across the city; once this factor is considered, the benefits of location-specific pricing appear to be relatively modest.

A second observation is that the results are higher in NYC than they are in LA (Figure 6, all three panels). This is because traffic patterns in NYC are more uniform than they are in LA. More specifically, if the components of σ vary too much, then we say the traffic in that city is relatively non-uniform as some locations are significantly more popular than others. A similar argument applies if the components of the distance vector $\mathbf{d}$ vary significantly. In our model the equilibrium prices and commissions depend on σ_i and d_i; if they vary too much across locations, then so should prices and commissions. A non-uniform traffic structure, therefore, calls for varied prices and commission rates. The implication is that in a city where these variables differ significantly, pursuing a non-flexible policy is more costly. In NYC the coefficient of variation for σ is 0.98, and for $\mathbf{d}$ it is 0.32. The corresponding numbers in LA are 1.52 and 0.35, implying that LA indeed has a more varied traffic structure than NYC. This explains why in the simulations the inflexible rules fare worse in LA than they do in NYC.

To confirm these insights, we randomly generated 100 cities, each consisting of 20 locations with distances varying from 4 to 12 miles. Accompanying transition matrices, too, were randomly generated. In each map, we computed the profits under each scheme, as well as the coefficient of

The resulting traffic flows revealed a highly similar traffic structure to that in our scraped Uber dataset. In both cases, core Manhattan neighborhoods dominate overall ride activity. The yellow cab data, for instance, shows Upper East Side (North and South), Midtown, Lincoln Square, Upper West Side, and Chelsea as the most active locations. These closely overlap with the top-ranked zones in the Uber data. Although the data come from different providers and are nearly a decade apart, the similarity in spatial patterns supports the reliability of our scraped data as a representation of NYC traffic flows. Due to their similar patterns and for brevity, we did not include a separate figure for the yellow cab data. A similar comparison was unfortunately not feasible for Los Angeles, due to the lack of publicly available ride-level data of comparable quality.

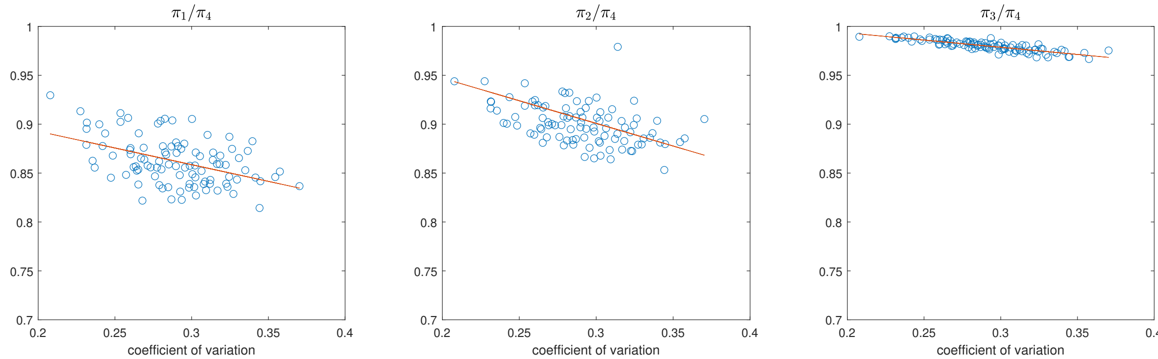

Figure 7: Traffic imbalance and profits in randomly generated cities

variation of σ: the higher the coefficient of variation, the more varied the traffic pattern in that city. We, then, plotted the profit ratios $\frac{\pi_1}{\pi_4}$, $\frac{\pi_2}{\pi_4}$ and $\frac{\pi_3}{\pi_4}$ in each map against the corresponding coefficient of variation and observed a clear downward trend in each panel in Figure 7. This confirms the previous insight that following a fixed rule—price-wise, commission-wise or both—becomes more costly for the platform as traffic patterns become more non-uniform (proxied by the coefficient of variation).

Social Welfare. Figure 8 plots social welfare, defined as the sum of platform profits, consumer surplus, and driver surplus, for both cities[24]. The patterns closely mirror those observed in the profit simulations: flexible commission models (3 and 4) deliver the highest levels of social welfare, while fixed commission models (1 and 2) consistently under-perform.

The reason is straightforward. Models 3 and 4 make the most efficient use of the driver fleet: they avoid bottlenecks, fully utilize available drivers, and achieve more matches than fixed commission models. This efficiency allows them to operate with a smaller fleet size, which in turn raises platform profits and consumer surplus. The only trade-off is that a smaller fleet generates less driver surplus, since relatively fewer drivers are earning above their participation thresholds. As a result, the performance gap in terms of social welfare is somewhat narrower than the gap in profits, but the ranking of the models remains unchanged.

Overall, the inclusion of driver surplus does not alter the main conclusion: flexible commission models remain the dominant performers, and their advantage persists when all stakeholders' interests are taken into account.

6 Conclusion

On-demand platforms are characterized by the flexible nature of work and the supply of independent workers, who have significant discretion over when and where to work, whether to continue working for a given platform or switch to an alternative work opportunity. Platforms, therefore, need to provide attractive earning opportunities and devise effective compensation mechanisms to incentivize and retain these independent workers. In addition, a key feature and complexity of ride-sharing platforms is the spatial differentiation of supply and demand with varying earning opportunities,

[24]For each value of μ, we determine the optimal entry level θ^* under each operating model along with the corresponding equilibrium compensation w^*. In the labor supply curve, θ^* marks the marginal driver, who is indifferent between participating at w^* and staying out. Drivers below this margin would have been willing to enter at lower pay, so the difference between their participation thresholds and w^* represents an additional rent. Summing these rents across all such drivers yields the total driver surplus.

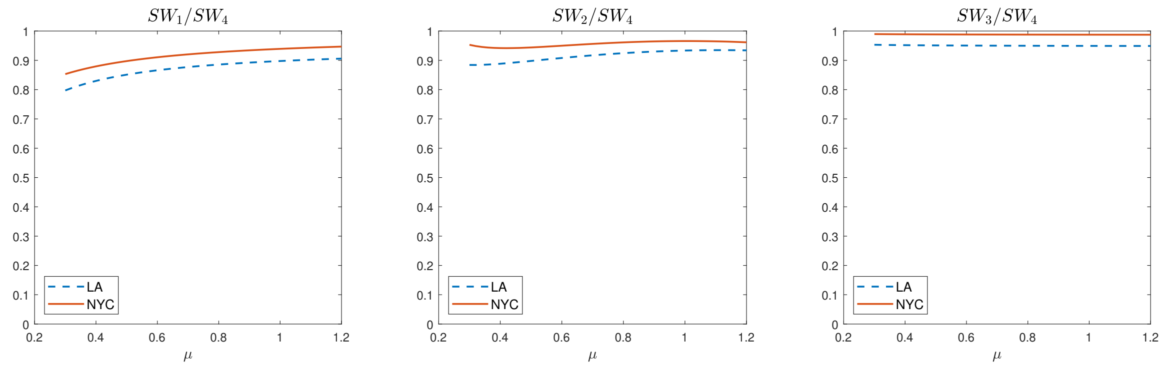

Figure 8: Social Welfare

which may lead drivers to concentrate in high-demand areas and leaving other areas with reduced service.

In light of these considerations, the platform should adopt an operational framework that explicitly accounts for the inter-dependencies among pricing strategies, consumer demand, driver entry and exit, and drivers' search behavior across locations. To sustain an optimal fleet size and ensure consistent service to customers, the platform must design compensation structures that align driver incentives with market conditions and are tailored to the specific features of each market.

Our analysis reveals several new insights and offers actionable operational strategies to the platforms. The results from our analytically tractable model highlight critical advantages of a flexible (location-specific) commission policy, which we believe has significant practical potential to be implemented for operational advantage. Indeed, ride-sharing platforms increasingly explore various ways to address location-specific supply and demand imbalances. For example, Lyft has introduced targeted promotions and incentives in Bonus Zones—areas with a high demand and a low number of drivers— while Uber has implemented Boost+ zones, which pay drivers extra for trips that begin in designated areas (Lyft, 2024; Uber, 2022a). Our study offers a systematic and strategic understanding of why flexible commission policy is an effective tool to address such imbalances.

In the absence of flexible commissions, the platform needs to address the supply and demand imbalances through price interventions, which suppress consumer demand and do not suitably incentivize drivers. In contrast, flexible commission policies can efficiently utilize available vehicles without resorting to unnecessary price hikes. Moreover, in a setting with free entry, such efficient utilization of drivers reduces the need for an excessively large fleet. Even though a smaller fleet size is associated with smaller driver surplus, flexible commission models still lead to more matches, resulting in higher platform profits, increased consumer surplus and ultimately higher social welfare.

A second advantage of flexible commission policies is their role in driver retention. When drivers struggle to find rides, they may switch to a different platform or exit the market altogether (Hall and Krueger, 2018; Allon et al., 2023a,b). By allowing the platform to better allocate drivers across locations, such policies increase the likelihood that each driver secures a match. This improved utilization lowers the risk of early market exit, resulting in reduced turnover and stronger retention. This represents a significant operational advantage for the platform.

Ride-sharing platforms are paying closer attention to local labor market conditions (Uber, 2022b), yet among various alternative pricing policies and implementation options, it is not clear which strategies would work better and why. Our analysis offers a clear managerial insight in response to labor market fluctuations: when drivers become less responsive to earning opportunities, the platform should prioritize adjusting commissions rather than prices. Raising commissions improves driver participation without dampening consumer demand, whereas raising prices risks reducing demand without effectively addressing the labor supply challenges.

Finally, while flexible (location-specific) pricing may be attractive for the platform, a well-documented concern in practice is the negative behavioral reaction by some customers who perceive flexible pricing as "unfair" and "exploitative", akin to customer reaction to surge pricing (Lowrey, 2014; Dholakia, 2015; Morrow, 2024). With this in mind, a noteworthy result of our analysis is that fixed pricing can, in fact, outperform flexible pricing. This finding highlights a critical trade-off for ride-sharing platforms and offers an important actionable insight. The behavioral reaction of fairness-sensitive customers can reduce the efficiency benefits of adopting flexible pricing. As such, if the percentage of such customers is large enough, then using a simple fixed pricing policy (coupled with flexible commissions) is a more effective and profitable tool in comparison to flexible pricing.

Our real-world calibration of the model for New York City and Los Angeles, and subsequent simulations in randomly generated cities, confirm our results and provide an additional managerial insight. We find that the performance of operating models critically depends on how balanced a city's traffic structure is in terms of trip lengths and traffic flows. If these parameters show significant variation, then pursuing a non-flexible policy is more "costly" for the platform. Thus, a key takeaway for platform managers is that cities with more uneven traffic patterns stand to gain the most from adopting flexible policies.

Our study comes with some limitations. We acknowledge that our model is stylized and some of our findings are descriptive in nature; as such, their implementation in the field requires a careful approach. However, we believe that our analytical model complements conventional big data analysis by offering a structured framework to uncover the underlying mechanisms and strategic interactions among ride-hailing stakeholders—drivers, customers, and the platform itself. Indeed, while platforms collect vast amounts of transactional data, big data models often function as black boxes, identifying correlations or making predictions, but not fully explaining why certain pricing and commission strategies work better than others. Moreover, since big data analytics inherently relies on observed outcomes, it may struggle to evaluate untested interventions or policy changes—such as how driver retention or the composition of the driver workforce might change under a redesigned pricing or commission policy.

These limitations highlight the need for a theory-driven framework that clarifies how platform policies shape outcomes across the system. Our model offers a general equilibrium approach that captures the effects of pricing and commission strategies on key performance dimensions. It generates testable hypotheses about how these policies influence driver entry and retention, profitability, and overall platform efficiency—insights that can support empirical research and inform managerial experimentation.

While the model is stylized, it serves as a tool to anticipate the effects of platform policies in environments where empirical evidence may be limited or unavailable. The value of the model lies in its ability to inform, not replace, data-driven decision-making. Of course, the effectiveness of any specific intervention and policy examined in our study ultimately depends on empirical validation and real-world experimentation.

References

AInvest (2025). Commission cuts for top drivers. Retrieved from: ainvest.com/news/bolt-pilots-20-commission-cuts-top-drivers-lagos-boost-earnings-2507.

Allon, G., Cohen, M. C., Moon, K., and Sinchaisri, W. P. (2023a). Managing multihoming workers in the gig economy. *The Wharton School Research Paper*. Available at SSRN: ssrn.com/abstract=4502968.

Allon, G., Cohen, M. C., and Sinchaisri, W. P. (2023b). The Impact of Behavioral and Economic Drivers on Gig Economy Workers. *Manufacturing & Service Operations Management*, 25(4):1376–1393.

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

Bai, J., So, K. C., Tang, C. S., Chen, X., and Wang, H. (2019). Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*, 21(3):556–570.

Banerjee, S., Johari, R., and Riquelme, C. (2015). Pricing in ride-sharing platforms: A queueing-theoretic approach. In *Proceedings of the 16th ACM Conference on Economics and Computation*, pages 639–639.

Benjaafar, S. and Hu, M. (2020). Operations management in the age of the sharing economy: what is old and what is new? *Manufacturing & Service Operations Management*, 22(1):93–101.

Bhargava, H. K. and Rubel, O. (2019). Sales force compensation design for two-sided market platforms. *Journal of Marketing Research*, 56(4):666–678.

Bimpikis, K., Candogan, O., and Saban, D. (2019). Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769.

Bolton, L. E., Warlop, L., and Alba, J. W. (2003). Consumer Perceptions of Price (Un)Fairness. *Journal of Consumer Research*, 29(4):474–491.

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Brown, E. (2019). Uber and Lyft Face Hurdle of Finding and Keeping Drivers. *The Wall Street Journal*. wsj.com/articles/uber-and-lyft-face-tough-test-of-finding-and-keeping-drivers-11557673863.

Buchholz, N. (2022). Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *The Review of Economic Studies*, 89(2):556–591.

Bursztynsky, J. (2021). Why many Uber and Lyft drivers aren't coming back. *CNBC*. cnbc.com/2021/07/04/why-many-uber-and-lyft-drivers-arent-coming-back.html.

Cachon, G. P., Daniels, K. M., and Lobel, R. (2017). The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*, 19(3):368–384.

Castillo, J. C. (2023). Who Benefits from Surge Pricing? *SSRN 3245533*.

Castillo, J. C., Knoepfle, D., and Weyl, G. (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242. ACM.

CBS News (2024). Lyft's Price Lock: A $2.99 Monthly Plan to Avoid Surge Pricing. cbsnews.com/news/lyft-monthly-membership-surge-pricing.

Chakravarty, A. K. (2021). Blending capacity on a rideshare platform: Independent and dedicated drivers. *Production and Operations Management*, 30(8):2522–2546.

Chan, T. Y., Li, J., and Pierce, L. (2014). Compensation and peer effects in competing sales teams. *Management Science*, 60(8):1965–1984.

Chen, M. K. and Sheldon, M. (2015). Dynamic pricing in a labor market: Surge pricing and the supply of Uber driver-partners. *Working Paper. (UCLA Anderson School)*.

Cohen, M. C. and Zhang, R. (2022). Competition and coopetition for two-sided platforms. *Production and Operations Management*, 31(5):1997–2014.

Cook, C., Diamond, R., Hall, J. V., List, J. A., and Oyer, P. (2020). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies*.

Cramer, J. and Krueger, A. B. (2016). Disruptive change in the taxi business: The case of Uber. *American Economic Review*, 106(5):177–82.

Dholakia, U. (2015). Everyone hates Uber's surge pricing. Here's how to fix it. *Harvard Business Review*. hbr.org/2015/12/everyone-hates-ubers-surge-pricing-heres-how-to-fix-it/.

Einav, L., Farronato, C., and Levin, J. (2016). Peer-to-peer markets. *Annual Review of Economics*, 8:615–635.

Eisenmann, T., Parker, G., and Van Alstyne, M. W. (2006). Strategies for two-sided markets. *Harvard Business Review*, 84(10):92.

Eyster, E., Madarász, K., and Michaillat, P. (2021). Pricing Under Fairness Concerns. *Journal of the European Economic Association*, 19(3):1853–1898.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Grinstead, C. M. and Snell, J. L. (1998). *Introduction to Probability*. American Mathematical Society.

Guda, H. and Subramanian, U. (2019). Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, 65(5):1995–2014.

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.

Hu, M. and Zhou, Y. (2019). Price, wage and fixed commission in on-demand matching. *Working Paper*. ssrn.com/abstract=2949513.

Huet, E. (2015). Uber's Ever-Renewing Workforce: One-Fourth Of Its Current U.S. Drivers Joined Last Month. *Forbes Magazine*.

Jain, S. (2012). Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Science*, 31(5):855–869.

Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review*, 76(4):728–741.

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

Lowrey, A. (2014). Is Uber's Surge-Pricing an Example of High-Tech Gouging? *The New York Times*. nytimes.com/2014/01/12/magazine/is-ubers-surge-pricing-an-example-of-high-tech-gouging.html.

Luckerson, V. (2014). Lyft Is Finally Launching in New York, Will Take on Uber. Retrieved from: https://time.com/2967215/lyft-new-york-uber/.

Lyft (2024). Bonus zones. From: help.lyft.com/hc/et/driver/articles/6198177189-driver-bonus-zones.

Mahdawi, A. (2018). Is your friend getting a cheaper Uber fare than you are? *The Guardian*. theguardian.com/commentisfree/2018/apr/13/uber-lyft-prices-personalized-data.

Morrow, A. (2024). Why 'dynamic' pricing feels like such a scam. *CNN*. (April 14), edition.cnn.com/2024/04/03/business/dynamic-surge-pricing-nightcap/index.html.

Naumov, S. and Keith, D. (2022). Optimizing the economic and environmental benefits of ride-hailing and pooling. *Production and Operations Management*.

OutlookBusiness (2025). Zero commission model. Retrieved from: outlookbusiness.com/start-up/news/ola-extends-zero-commission-model-to-cab-drivers-after-autos.

Parker, G. G. and Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 51(10):1494–1504.

Paul, K. (2021). Uber and Lyft drivers join day-long strike over working conditions. *The Guardian*. theguardian.com/technology/2021/jul/21/uber-lyft-drivers-strike-app-based-work-gig-economy.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029.

Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3):645–667.

Schöttner, A. (2017). Optimal sales force compensation in dynamic settings: Commissions vs. bonuses. *Management Science*, 63(5):1529–1544.

Schuitema, G., Steg, L., and van Kruining, M. (2011). When Are Transport Pricing Policies Fair and Acceptable? *Social Justice Research*, 24(1):66–84.

Tan, B., Anderson Jr, E. G., and Parker, G. G. (2020). Platform pricing and investment to drive third-party value creation in two-sided networks. *Information Systems Research*, 31(1):217–239.

Uber (2019). Uber Newsroom. (last accessed: April 14, 2019) Available at www.newsroom.uber.com/wp-content/uploads/2014/07/uberlaconnectome.

Uber (2020). Uber service fee. Available at uber.com/gh/en/drive/basics/tracking-your-earnings.

Uber (2022a). Introducing Boost+. At www.uber.com/blog/introducing-boost.

Uber (2022b). Uber boosts driver rates to help meet growing demand. At uber.com/en-GB/newsroom/uber-boosts-driver-rates-to-help-meet-growing-demand/ (last accessed date: April 3, 2024).

Uber (2024). Uber announces results for first quarter 2024. Available at investor.uber.com/news-events/news/press-release-details/2024/Uber-Announces-Results-for-First-Quarter-2024.

Wang, Y., Wu, C., and Zhu, T. (2019). Mobile hailing technology and taxi driving behaviors. *Marketing Science*, 38(5):734–755.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *American Economic Review*, 100(4):1642–1672.

Wohlsen, M. (2013). Uber boss says surging prices rescue people from the snow. *Wired Magazine*. wired.com/2013/12/uber-surge-pricing/.

Xia, L., Monroe, K. B., and Cox, J. L. (2004). The Price is Unfair! A Conceptual Framework of Price Fairness Perceptions. *Journal of Marketing*, 68(4):1–15.