Congestion Charges Under Market Power: An application to ride hailing in Bogotá, Colombia

Julian A. Gomez-Gelvez*

November 27, 2020

JOB MARKET PAPER

Download the most recent version

Abstract

The rapid growth of ride-hailing services over the last decade threatens to worsen traffic congestion in cities worldwide. Economists usually prescribe a Pigouvian tax or congestion charge, equal in size to the marginal external cost of congestion, to treat this illness and contain excessive growth. However, ride-hailing markets suffer from another imperfection. They are usually concentrated in the hands of very few digital platforms like Uber. Platforms can then exert market power and raise prices above competitive levels. Under these two conditions (negative externalities and market power), the optimal congestion charge is smaller than the marginal external cost and may even turn negative. In this paper, I build a structural model of ride hailing to compute the optimal congestion charge for a ride-hailing market monopolized by a digital platform. I calibrate the model to the morning peak period in Bogotá, Colombia, in 2019 and find that the markup imposed by the monopolist platform (in the form of a gap between the prices charged to riders and paid to drivers, usually known as platform commission) covers about 70% of the marginal external congestion cost of ride hailing. As a result, the optimal congestion charge corresponds to only 40% of this external cost. The optimal charge takes into account that the platform adjusts prices in response to the charge, which causes an incomplete pass-through of the charge to riders.

*University of Maryland, College Park, Department of Agricultural and Resource Economics. Email: jgomezge@umd.edu. I would like to thank Rob Williams, Stephen Salant, Joshua Linn and seminar participants at the World Congress of Environmental and Resource Economics (WCERE, 2018), the International Transportation Economics Association (ITEA, 2019) and the Summer Conference of the Association of Environmental and Resource Economists (AERE, 2019) for useful comments and suggestions.

1 Introduction

The rapid growth of ride-hailing services mediated by digital platforms like Uber, Lyft or DiDi over the last decade has caused concerns about their potential to worsen traffic congestion and other transportation-related externalities such as air and climate pollution. As a result, congestion charging (the adaptation of a Pigouvian tax to traffic congestion externalities) has been suggested as a mitigation strategy. For example, New York City implemented in February 2019 a congestion surcharge of $2.75 for ride-hailing trips entering or passing through Manhattan south of 96th street (New York City Taxi and Limousine Commission, 2019).[1]

However, negative externalities are not the only failure of ride-hailing markets. Market power is also alive. Even though there are usually thousands of riders and drivers, only one or two digital platforms generally control the market in each city. For example, Uber and Lyft control almost the entire market in most U.S. cities (Statista, 2020). In their typical business model, ride-hailing platforms not only match riders and drivers but also set prices to both sides of the market (charged to riders and paid to drivers). Platforms can then exert market power to impose a profit-maximizing gap between these two prices (commonly known as platform commission), in the same manner a monopolist producer imposes a profit-maximizing markup. Ride-hailing markets tend to concentrate in very few platforms mainly because of network effects in wait times. A platform serving proportionally more riders and drivers can offer lower wait times to travelers (Frechette, Lizzeri, & Salz, 2019).

Economists have long recognized that market structure can significantly influence the efficiency of Pigouvian taxes (Buchanan, 1969). This insight can be easily grasped. Picture a market monopolized by a producer, who imposes a profit-maximizing markup and causes an external cost (larger than the markup). A regulator concerned about the externality may wish to impose a Pigouvian tax equal to the marginal external cost (MEC), but this policy would push the price faced by consumers above the optimal level. The optimal tax to address the externality must be smaller than the MEC because there is a smaller gap to bridge. In a more extreme, but equally plausible, scenario, the markup may be larger than the MEC, in which case *any* positive tax would be detrimental. In a lucky situation, MEC=markup, the

[1]New York is also expected to be the first U.S. city to introduce congestion pricing for private vehicles in 2021. Even though economists have long argued for congestion pricing as a tool to manage traffic congestion, only a handful of cities around the world have actually implemented it. The New York case suggests that it may be easier to implement congestion pricing for ride hailing than for private cars. At least on the technological side, barriers are substantially lower because platforms already have the technology in place to identify and charge individual trips, so no need for E-Z passes or cameras taking photos of license plates. Other cities that have implemented surcharges on ride hailing include Mexico City (1.5%), Chicago (US$0.69), Rio de Janeiro (1%), Calgary (CAN$0.30) and San Francisco (3.25%) (Yanocha & Mason, 2019).

monopolist charges a socially efficient price and there is no need for government intervention to achieve efficiency.

In this paper, I apply the previous observation to ride-hailing markets in order to judge the merit of a congestion charge. Since ride-hailing markets feature both externalities and market power, the optimal congestion charge is smaller than the MEC of congestion, and may actually turn negative if the markup is substantial. This paper provides the first empirical comparison between congestion externalities and market power in the ride-hailing industry. To do so, I set up and calibrate for Bogotá, Colombia, a structural model of ride hailing, which ultimately allows me to estimate the optimal charge under a monopolistic market structure.

The structural model has four components. The first one is a demand model for ride-hailing services. I propose and estimate empirically based on stated-preference surveys a demand model that allows for individual heterogeneity in reservation values and values of time. Importantly, the model allows also for correlation between these two dimensions of heterogeneity. This correlation, usually neglected in transportation demand models, can have a significant impact on the divergence between monopolistic and optimal price levels (Mills, 1981). Travel time for ride hailing is the sum of wait time and in-vehicle time. Both of these time factors depend on the number of ride-hailing vehicles on the street. In-vehicle time increases with the number of vehicles due to traffic congestion. Wait time decreases with it because there is a better chance an idle vehicle is close to the rider's location. As a result, the number of travelers willing to hail a ride depends not only on the price charged by the platform, but also on the number of ride-hailing vehicles available for service.

The second component is a supply of drivers that adjusts to achieve a fixed revenue per hour. The assumption behind a constant revenue per hour is that all potential drivers have the same reservation wage, which includes vehicle operating expenses (e.g. gasoline) and net earnings. J. V. Hall, Horton, and Knoepfle (2020) provide empirical evidence suggesting this assumption is a good approximation to the labor supply of ride-hailing drivers. However, it brings two important and interrelated implications, which should be kept in mind to ponder some of the results. First, no surplus is created on the side of drivers (all drivers earn their reservation wage). Second, the monopsonistic position of a platform in the market for ride-hailing drivers does not result in a markdown in the price paid to drivers. The uniform-reservation-wage assumption is equivalent to a constant marginal cost of production.

The third component is a matching process between riders and drivers, which determines wait times. I assume platforms match riders to the closest idle vehicle, and idle vehicles are evenly distributed over the service area. These two assumptions determine a mathematical relationship, first derived by Arnott (1996), between average or expected wait time and the density of idle vehicles in the service area. The number of idle vehicles is, in steady state, a function of the total number of vehicles available for service, the number of trips requested per hour and average travel time. This function introduces network effects into the picture, because a proportional increase in riders and drivers (i.e. an increase in the platform's scale) raises the number of idle vehicles and consequently lowers wait times.

The fourth and last component of the structural model is an empirical estimate of the marginal effect of additional vehicles on average travel speed. This magnitude is clearly at the heart of congestion externalities. I measure it for Bogotá based on data from its 2019 Mobility Survey and applying the methodology originally proposed by P. A. Akbar and Duranton (2017). In this methodology, the effect of additional vehicles on travel speed is identified from changes in traffic volume throughout the day, controlling for concurrent changes in trip and traveler characteristics.

These four components interact to determine the number of riders and drivers in equilibrium for a given set of prices per trip charged to riders and paid to drivers. Once prices are set, ride-hailing markets clear on travel time (in-vehicle plus wait).

In the base scenario, a monopolistic platform sets prices to maximize profit. Profit equals the product between the number of riders and the price gap or platform commission. In an ideal scenario, a benevolent social planner controls the platform and sets prices to maximize total welfare. Total welfare includes the surplus created for ride hailers (there is no surplus for drivers), minus the cost of vehicles and drivers and the traffic congestion externality imposed on other road users. Finally, in an economist's dream, a private platform sets prices but a regulator is able to force socially efficient outcomes through taxation.

I derive analytical conditions for the profit- and welfare-maximizing price gaps. These conditions reproduce the contest between markup and marginal external cost that drives the sign and magnitude of the optimal congestion charge. Additionally, they reveal that a private platform internalizes (with a distortion) the congestion externality each ride hailer imposes on other ride hailers (not on other road users). As a result, the platform internalizes a larger portion of the total external cost of congestion as ride-hailing vehicles become a larger percentage of total traffic volume, potentially weakening the motivation for a congestion charge.

I also calibrate the components of the structural model to the morning peak period of an average weekday in 2019 in Bogotá, Colombia. Bogotá is a highly dense and congested city of about 7.5 million inhabitants.[2] According to its 2019 Mobility Survey (Secretaría Distrital de Movilidad, Bogotá D.C., 2019), the number of ride-hailing trips per hour during the morning peak period was about 11,000, which is a high number but represents only 1.1%

[2] Bogotá usually tops worldwide congestion rankings such as the Global Traffic Scorecard (Inrix, 2019).

of all trips in the city.[3] The survey also reveals that Uber had a strong hold on the market with about 70% of all ride-hailing trips.

The numerical results for Bogotá indicate that, in a monopolistic scenario, the platform imposes a 20.1% gap or commission between the prices charged to riders and paid to drivers. This gap represents the markup imposed by the platform. The rider fare in this scenario is only about 5% larger than the average fare observed in the mobility survey, which suggests the monopolistic scenario is a good approximation to the situation of Bogotá's ride-hailing market in 2019. When a social planner manages the platform, the price gap increases to 26.4%. This optimal gap reflects mainly the external cost ride hailing imposes on other road users through traffic congestion. In this case, the marginal external cost turns out to be larger than the markup.

The government of Bogotá does not have to take control of the platform in order to force socially efficient outcomes. Regulators can realize about 95% of the welfare gains available from the unregulated scenario by imposing an appropriate tax (or congestion charge) to riders' fare. Even though regulators should in theory take care of both sides of the market (riders and drivers) in order to fully achieve socially efficient outcomes, an appropriate tax on the side of riders does most of the job. The optimal congestion charge does not equal the difference between the monopolistic and optimal price gaps because the private platform lowers the price charged to riders (net of tax) as a response to the tax. The correct charge must then be larger than the initial price-gap difference to account for the incomplete pass-through. The model reveals a monopolistic pass-through of 0.77 (i.e. in response to a $1,000 tax, the platform lowers its price by about $230).

The optimal congestion charge is COL$1,390 for an average-distance trip during the morning peak period. This charge represents about 12% of the current rider fare, or COL$180 per kilometer.[4] It also represents only 40% of the marginal external cost of congestion. The optimal tax is smaller than the marginal external cost due to the monopolistic structure of the market.

A few considerations about these results are in order. First, the charge does not account for other transportation-related externalities such as traffic accidents and air pollution, which would enlarge it.[5] Second, the optimal charge for other time periods is likely to vary signifi-

[3]As a comparison, the number of ride-hailing trips per hour during the evening peak in San Francisco in 2016 was about 14,000, while ride-hailing trips accounted for 9% of all weekday trips in the city (San Francisco County Transportation Authority, 2017).

[4]The optimal congestion charge in the model is proportional to trip distance, so it should be applied on a per-kilometer basis or as a percentage of the fare, not as a fixed amount to all trips.

[5]Congestion externalities are usually larger than accidents and air pollution (see for instance Parry and Small (2005)), especially during peak hours, so the optimal charge may not increase considerably due to these other externalities.

cantly, especially for periods with low traffic volume. Finally, this paper assumes regulators do not impose a congestion charge on private cars. Such a (desirable) policy would affect the optimal charge on ride hailing to the extent it reduces traffic volume or increases the demand for ride hailing.[6]

The remainder of this paper contains four sections. Section 2 details the components of the structural model and presents analytical conditions for the profit- and welfare-maximizing price gaps. Section 3 describes the data and empirical estimations carried out to calibrate the model to Bogotá. Section 4 reveals the main numerical results. I wrap up in Section 5 with the main conclusions. Before all that, I will describe how this paper connects and contributes to several bodies of literature.

Related literature

The emergence of ride-hailing platforms over the last decade has inspired a rapidly growing literature with diverse emphases. One initial concern has been the potential of ride hailing to increase vehicle-miles traveled (VMT) in cities, consequently exacerbating traffic congestion and other transportation-related externalities. There was an initial debate in the transportation literature about whether ride hailing contributes or not to urban VMT. Platforms regularly argue that ride hailing can reduce VMT by facilitating access to public transportation, reducing the need to own cars or providing shared services. A few studies supported these arguments. For instance, J. D. Hall, Palsson, and Price (2018) found that Uber increased transit ridership by 5% in average two years after entry to U.S cities. However, the most recent literature with detailed data concludes that ride hailing does contribute to VMT and congestion, mainly because it replaces many trips that would have been made by more sustainable modes such as public transportation. For example, Erhardt et al. (2019) conclude that ride hailing had a significant effect on congestion in San Francisco between 2010 and 2016. Tirachini (2020) provides an international review including experiences from developing countries.

On the other hand, several studies have quantified the economic value created by ride-hailing platforms. Cohen, Hahn, Hall, Levitt, and Metcalfe (2016) and Lam and Liu (2017) estimate that riders gain a surplus of \$1.60 and \$0.72, respectively, per dollar spent on platforms in major U.S. cities. Frechette et al. (2019), Buchholz (2020), Bian (2018) and Shapiro (2018) develop dynamic and spatial equilibrium models to estimate the efficiency gains that electronic matching offers over street hailing. Castillo, Knoepfle, and Weyl (2018)

[6] Bogotá currently takes 50% of the private car fleet out of circulation during peak hours through a license plate-based restriction. This restriction probably contributes to the demand for ride hailing. Changes to this policy would also affect the optimal congestion charge on ride hailing.

and Castillo (2019) focus on the welfare gains available from dynamic pricing. Finally, Chen, Chevalier, Rossi, and Oehlsen (2019) measure the value drivers derive from being able to choose when to work.

This paper bridges these two opposing views of ride hailing. Even though digital platforms increased the efficiency of rider-driver matchings and implemented dynamic pricing, these improvements do not imply they raised overall welfare due to two factors. First, efficiency improvements may result in lower overall welfare in the presence of unregulated externalities such as traffic congestion. Second, the pricing strategies pursued by private platforms with market power may differ considerably from the socially optimal ones. Only through proper regulation we can extract the full benefits of technological improvements and guarantee they do not decrease welfare. The main objective of this paper is to design such regulation for the ride-hailing industry.

This paper also contributes to the literature on environmental regulation under market power. Buchanan (1969) pointed out that environmental regulation designed to completely internalize external damages in non-competitive industries may reduce welfare. Fowlie, Reguant, and Ryan (2016) confirm this possibility for the U.S. cement industry. They find that policies designed to internalize the social cost of carbon in this industry reduce welfare. Consequently, optimal carbon pricing involves firms only partially internalizing the social cost of carbon. My results offer a similar conclusion. The optimal congestion charge on ride hailing corresponds to only 40% of the marginal external cost of congestion.

Finally, this paper contributes to the economic literature on monopolistic pricing of congestible resources. Mills (1981) showed that a monopolist internalizes (with a distortion) the congestion effects on users of its resource, and revealed that the magnitude of this distortion depends on the correlation between reservation values and sensitivities to congestion (values of time in our context) in the population of potential users. Even though the transportation economics literature recognizes this effect (Brueckner, 2002; Verhoef & Small, 2004), this paper presents, to my knowledge, the first estimate of correlation between reservation values and values of time across a population of travelers.

2 Model and analytical solutions

This section details the four components of a model that intends to capture the main features of the ride-hailing industry as it has come to be in the last decade. Ride-hailing platforms electronically match riders and drivers, and set prices to both sides of the market.[7] Potential

[7]Other business models for platforms exist but have not become mainstream. For example, InDriver allows drivers to name their price for each trip while riders select the best offer. Empower does not impose

riders can check the price and estimates of in-vehicle and wait times on their smartphones before deciding to request a ride. On their side, drivers are free to choose when to be available for service, basing their decisions on previous experience about average earnings.[8] An economic model of ride hailing must then include at least three components: a demand model that represents travelers' alternatives, a supply model that represents drivers' decisions to work, and a matching process that determines wait times. Since one of the main objectives of this paper is to evaluate the impact of ride hailing on traffic congestion, I add a fourth component that endogenizes in-vehicle travel time as a function of the number of ride-hailing vehicles on the street.

I simplify platforms' pricing decisions down to one price to charge riders and one price to pay drivers for an average-distance trip. It is implicitly assumed that prices for diverse trips are proportional to distance, both in monopolistic and socially optimal scenarios.[9] Accordingly, the optimal congestion charge will be computed for an average-distance trip, but should be applied to diverse trips on a per-kilometer basis or as a percentage of the price charged by the private platform. The model evaluates pricing within a single time period of the day, ignoring any potential interdependencies in pricing across different time periods.[10]

2.1 Demand: Riders' side

The first time you install a ride-hailing app on your phone, you may decide to use it for a trip that you would have otherwise done by other means (e.g. by bus), or you may actually decide to travel because the app now makes it convenient. The population of potential ride hailers is then composed of people currently traveling by other modes as well as people not traveling. I will, nonetheless, most of the time refer to potential ride hailers as travelers to

a platform commission, but charges drivers a flat monthly fee.

[8]Drivers usually do not know the amount they will earn for, and have in general limited information about, a trip before accepting it. Therefore, their role in picking trips is small.

[9]The model does not miss an important discrepancy between monopolistic and optimal pricing to the extent that prices are in fact roughly proportional to distance in both scenarios. Profit-maximizing prices may not be proportional to distance if some traveler's characteristics, such as income, are significantly correlated with distance, because platforms may then price-discriminate based on distance (e.g. price long trips relatively higher if they are mainly made by high-income travelers). However, the usual pricing strategy by platforms, which includes a fixed charge plus variable charges based on distance and travel time, implies that prices are approximately proportional to distance. Socially optimal prices may not be proportional to distance if, for example, long trips tend to add proportionally more or less to congestion than short ones.

[10]Platforms may set prices for different time periods in an interdependent way if demand or supply are related across time periods. Demand may be related across time periods if travelers can adjust the start time of their trip in response to prices. Supply may also be related if drivers' decision to work in a time period depends on whether they work on an adjacent period (i.e. drivers may prefer to work consecutive rather than alternative periods). The time period of analysis must then be ample enough to believe it would be too costly for travelers to adjust their start times to other periods, and to believe drivers can choose to work during this period without having to work on adjacent ones.

ease wording. I will refer to people actually requesting rides as rider or rider hailers.

Travelers choose between ride hailing and their best alternative or outside option (other mode or not traveling) within the time period of analysis.[11] Normalizing the value of the outside option to zero, the value traveler i gets from ride hailing can be expressed as:

$$\text{Ride hailing} : V_i - \beta_i \cdot [t(d) + w(x,d)] - p \tag{1}$$

$$\text{Outside option} : 0$$

V_i is the reservation value in monetary terms traveler i assigns to ride hailing in comparison to her outside option. V_i reflects individual preferences as well as the characteristics of outside options (e.g. price and travel time of alternative modes). β_i is the value of time of traveler i. It is a measure of how travel time lowers reservation values.[12] $t(d)$ is in-vehicle travel time as a function of the number of ride-hailing vehicles on the streets. In-vehicle time increases with the number of vehicles due to traffic congestion (see Section 2.4). $w(x,d)$ denotes average or expected wait time as a function of the number of riders or trips requested per hour (x) and the number of vehicles (d).[13] Wait time increases in x and decreases in d (see Section 2.3). Finally, p is the price per trip charged to riders.

A reservation value (V_i) and a value of time (β_i) identify each traveler. A bivariate probability density function $f(\beta, V)$ can then represent the distribution of individual preferences across the population of potential ride hailers. For a given price (p), in-vehicle time (t) and wait time (w), traveler i chooses ride hailing if $V_i - \beta_i \cdot (t + w) - p > 0$. For a given set (p, t, w), riders can then be identified in a two-dimensional graph, with β in the horizontal axis and V in the vertical one, as those above and to the left of a ray that goes through $(0, p)$ and has slope $t + w$. Figure 1 illustrates this characterization, where S denotes the support of the bivariate distribution $f(\beta, V)$.[14]

Since riders raise wait times, the equilibrium number of riders for a given price and number of vehicles must solve the following equation (where N represents the size of the

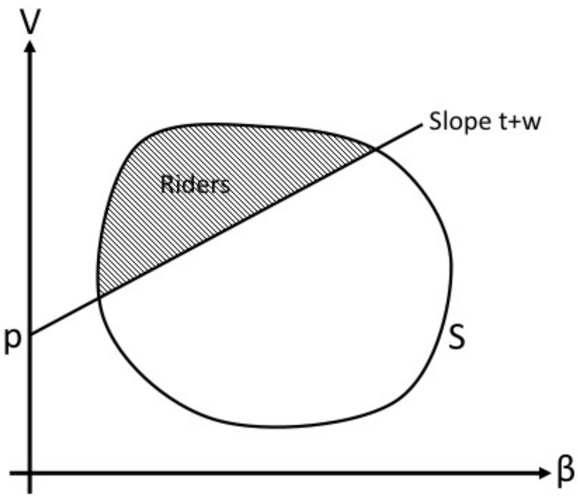
[11] I assume people make at most one trip during the time period of analysis. Not a ridiculous assumption, especially for peak periods.

[12] The transportation literature highlights that most travelers value wait time more than in-vehicle time (Small, 2012). I could allow these two valuations to vary independently across travelers, but that would introduce an additional dimension of individual heterogeneity, further complicating empirical estimation and numerical solutions.

[13] I assume one rider corresponds to one trip. The model does not consider shared rides.

[14] I assume the support lies entirely in the first quadrant (positive V and β). It would take a very strange individual to encounter a negative value of time. It is plausible for individuals to have negative reservation values, but their existence is immaterial for our purposes because they would never choose ride hailing, unless platforms decide to pay riders (negative price) or we figure out a way to travel back in time (negative travel time).

Figure 1: Riders for a given price (p), in-vehicle time (t) and wait time (w).



Notes: S identifies the support of the distribution of reservation values (V) and values of time (β) in the population of potential riders.

population of potential ride hailers):

$$N \int_{\beta=0}^{\infty} \int_{V=p+\beta(t(d)+w(x,d))}^{\infty} f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta = x \tag{2}$$

Equation 2 determines the number of riders (x) as an implicit demand function of price (p) and number of vehicles (d).

2.2 Supply: Drivers' side

The side of drivers is simpler. Drivers choose to work during the time period of analysis if expected earnings are higher than their reservation wage. Since drivers are usually responsible for all vehicle expenses (e.g. gasoline consumption), their reservation wage includes these expenses plus net earnings. I assume all potential drivers have the same reservation wage, denoted c (for cost, w was already taken by wait time). This assumption may seem extreme, but J. V. Hall et al. (2020) provide empirical evidence suggesting it is a good approximation to the labor supply of ride-hailing drivers. They show that after Uber-initiated price increases in U.S. cities, driver supply adjusted to bring hourly earnings back to their initial level.[15]

Expected hourly earnings for a driver equal the product of the number of trips per hour she expects to serve and the price per trip paid by the platform (q). Assuming trips are

[15]The uniform-reservation-wage assumption can also be considered an approximation to a large pool of potential drivers.

evenly distributed among drivers, the expected number of trips for a driver equals the ratio between total trips (x) and total drivers (d).

The equilibration process on the side of drivers is then straightforward. Drivers enter the market until expected earnings equal the common reservation wage (c). Mathematically:

$$c = \frac{x}{d} \cdot q \tag{3}$$

Equation 4 determines the number of drivers or vehicles (d) as a supply function of price (q) and number of riders per hour (x). Contrary to the demand function, the supply function can be made very explicit:

$$d = \frac{q}{c} \cdot x \tag{4}$$

The uniform-reservation-wage assumption brings two important and interrelated implications, which should be kept in mind to ponder some of the results. First, no surplus is created on the side of drivers (all drivers earn their reservation wage). Second, the monopsonistic position of a platform in the market for ride-hailing drivers does not result in a markdown in the price paid to drivers.

2.3 Matching and wait times

The way platforms match riders and drivers determines how long riders have to wait for their vehicles.[16] I assume platforms match riders to the closest idle driver, which is a natural assumption given that platforms have access to the location of riders and drivers. A driver (or vehicle) is idle if she is not currently busy serving a passenger or en route to pick one up.

Arnott (1996) showed that if I idle vehicles are evenly distributed over a service area of size A, so that the density of idle vehicles is $D = I/A$, the expected distance between a rider and the closest idle driver can be approximated as $\frac{1}{2\sqrt{D}}$. Assuming vehicles travel at speed v when en route to pick up a passenger, average or expected wait time (w) as a function of the density of idle vehicles is:

$$w(D) = \frac{1}{2v\sqrt{D}} \tag{5}$$

At any moment, the number of idle vehicles equals the total number of vehicles (d) minus the number of busy vehicles. In steady state, the expected number of busy vehicles equals the product of the number of trips per hour (x) and the average trip time ($t + w$).

[16] Wait time refers to the time it takes a driver to reach the location of the rider she has been matched with. Due to the stochastic nature of the process, there may be some additional wait time if there are no idle vehicles at the time the rider requests a ride. However, for the scale of ride hailing that concerns us (thousands of riders and drivers), this additional wait time is usually only a few seconds, so it is ignored. Li, Tavafoghi, Poolla, and Varaiya (2019) reach a similar conclusion for ride hailing in New York City.

Mathematically:

$$D = \frac{I}{A} = \frac{d - (t+w)x}{A} \tag{6}$$

Plugging this expression into Equation 5, expected wait time can be expressed as:

$$w = \frac{1}{2v\sqrt{\frac{d-(t+w)x}{A}}} \tag{7}$$

Equation 7 implicitly defines expected wait time as a function of the number of drivers and riders ($w(x,d)$ in Expression 1). This relationship, however, can be quite complex because w is in both sides of the equation and because in-vehicle time (t) and speed (v) are functions of the number of vehicles due to traffic congestion. Fortunately, we can get a good approximation assuming that speed and average trip time ($t+w$) are fixed.[17] Denoting the fixed average trip time as s (for service time), Equation 7 reduces to:

$$w(x,d) = \frac{1}{2v\sqrt{\frac{d-s\cdot x}{A}}} \tag{8}$$

Equation 8 now explicitly defines expected wait time as a function of the number of drivers and riders. This equation reveals network effects in the ride-hailing industry. If drivers and riders increase in the same proportion, expected wait time declines. For example, a platform with four times as many drivers and riders can offer half wait times in average. The effect is even more transparent when we explicitly take into account that the number of drivers adjust to the number of riders. Using Equation 4 to substitute d out of Equation 8, we can express expected wait time as a function of only the number of riders:

$$w(x) = \frac{1}{2v\sqrt{\frac{(q/c-s)x}{A}}} \tag{9}$$

Equation 9 reveals that expected wait time is a decreasing function of the number of riders.[18]

2.4 Traffic congestion and in-vehicle time

Additional vehicles on the streets, such as ride-hailing vehicles, tend to slow down all other vehicles and increase travel times. However, to be fair to ride hailing, we must consider

[17]This approximation is good under two conditions: (1) wait constitutes a small percentage of trip time, and (2) the effect of ride-hailing vehicles on travel speed is small in percentage terms. This approximation is used only for expositional purposes. I use Equation 7 to obtain numerical results.

[18]$q/c - s$ must be a positive quantity for an equilibrium with positive number of drivers and riders to exist (see Section 2.5).

what would happen if it were not available. As explained in Section 2.1, ride hailers would probably shift to a different mode of transportation or perhaps not travel. If the main alternative for ride hailers is, for example, not to travel or to bike, the effect of ride hailing on traffic congestion would be significant. On the other extreme, if their main alternative is to use private cars, the effect could be small or even negative.[19] The case of public modes such as buses and taxis probably lies somewhere in between depending on how the supply of public vehicles adjusts to changes in demand.[20]

Expression 1 introduced an in-vehicle travel time function ($t(d)$) that depends only on the number of ride-hailing vehicles on the street, which implies all other traffic, such as cars and buses, remains constant as the scale of ride hailing varies. There are two important assumptions behind this simplification. First, there is no substitution between ride hailing and cars. Second, the supply of public vehicles that use common streets (e.g. regular buses and taxis) does not adjust to changes in demand. Section 3.1 argues that these assumptions are plausible for Bogotá. However, they may not hold for other cities. Most importantly, if there is significant substitution between ride hailing and cars, the demand model introduced in Section 2.1 should be expanded to explicitly include private car as an option, while in-vehicle time should depend both on the number of ride-hailing vehicles and the number of private cars. These adjustments can have a significant effect on the size of the optimal congestion charge, because demand shifts from ride hailing to private cars are not likely to have a major impact on congestion.[21]

The external cost ride hailing imposes on other road users due to traffic congestion can be approximated as the product of the number of road users, their average value of time and the average in-vehicle time increment caused by ride-hailing vehicles. In turn, this time increment can be approximated as the marginal increase caused by one vehicle times the number of vehicles.[22] Mathematically:

$$EC(d) = \beta_{other} \cdot N_{other} \cdot Mg\_time \cdot d = MEC \cdot d \tag{10}$$

where $EC(d)$ is the total external congestion cost caused by ride-hailing vehicles, N_{other} denotes the number of other road users per hour during the time period of analysis, β_{other} represents their average value of time, and Mg_time corresponds to the marginal in-vehicle

[19]The net effect of ride hailing on congestion in comparison to private cars may be negative if private cars have to cruise for parking. On the other hand, ride-hailing vehicles usually deadhead in between trips.

[20]For example, if the number of buses is fixed, a shift of travelers from buses to ride hailing would increase congestion (although it would also benefit bus users due to reductions in crowding and possibly wait times).

[21]The optimal congestion charge on ride hailing tends to decrease as substitution with cars intensifies (i.e. as more ride hailers have private car as their best alternative).

[22]Since the relationship between traffic volume and travel time is in general non-linear, this last approximation is valid as long as the number of ride-hailing vehicles represents a small percentage of total traffic.

travel time increase caused by an additional vehicle on the street. $MEC = \beta_{other} \cdot N_{other} \cdot Mg\_time$ compiles the marginal external cost that one ride-hailing vehicle imposes on other road users through traffic congestion. Section 3.3 describes the empirical approach I use to estimate Mg_time for the morning peak period in Bogotá. I will also use this parameter to specify the in-vehicle travel time function for ride hailing ($t(d)$).

2.5 Equilibrium number of riders and drivers

Equation 2 implicitly defines the number of riders or trips per hour (x) as a function of the number of drivers or vehicles (d) and the price per trip charged to riders (p) (demand side). Similarly, Equation 4 determines, in a simpler and explicit way, the number of drivers as a function of the number of riders and the price per trip paid to drivers (q) (supply side). As you might have guessed, these two equations together establish equilibrium numbers of riders and drivers (x, d) for a given set of prices (p, q). This section shows there may be two equilibrium points for a given set of prices, but only one of them has positive numbers of riders and drivers.

Let's start by using Equation 3 to substitute x out of Equation 8 and express expected wait time as a function of only the number of drivers:

$$w(d) = \frac{1}{2v\sqrt{\frac{(1-c\cdot s/q)d}{A}}} \tag{11}$$

We can now insert this expression for expected wait time into Equation 2 to express the number of riders as a function of the number of drivers:

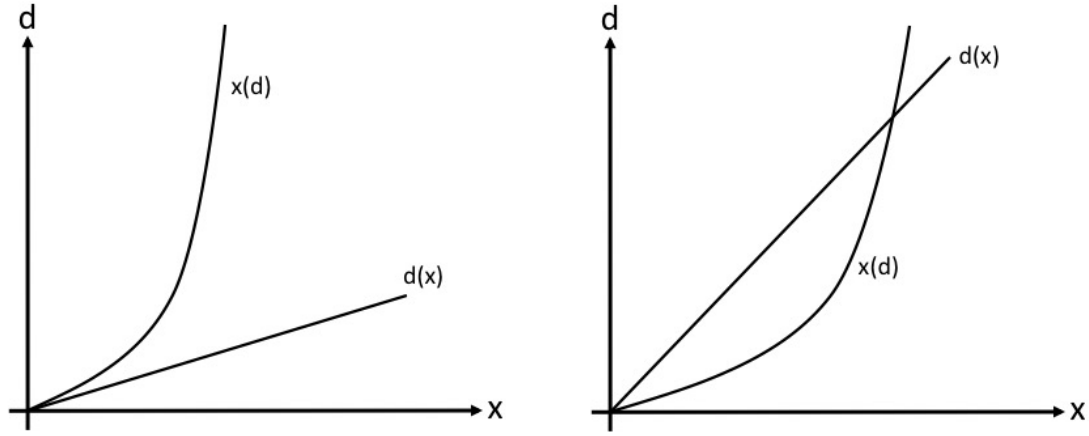
$$x(d) = N\int_0^\infty \int_{p+\beta(t(d)+w(d))}^\infty f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta \tag{12}$$

Recall also that Equation 4 determines the number of drivers as a function of the number of riders:

$$d(x) = \frac{q}{c}\cdot x \tag{13}$$

Equations 12 and 13 allow us to examine the equilibrium points in a two-dimensional graph with x in the horizontal axis and d in the vertical one (Figure 2). Equation 13 represents a straight line that goes through the origin, while Equation 12 describes a concave function that also goes through the origin and approaches a finite limit as d goes to infinity. According to the price levels (p, q) and the other parameters that influence Equations 12 and 13, the origin may be the only point where these two lines intersect (Figure 2a), so it may be the only equilibrium. More sensible pricing, however, can lead to the situation of Figure 2b, where there is an additional equilibrium with positive numbers of riders and drivers.[23]

Figure 2: Equilibrium number of riders (x) and drivers (d).



(a) One equilibrium point. (b) Two equilibrium points.

Notes: Panel (a) illustrates a situation where the only equilibrium has no drivers and no riders. In panel (b), there is an additional equilibrium with positive numbers of riders and drivers.

The fact that an equilibrium point without riders and drivers is always possible follows from the interdependency of supply and demand. Without riders, there is no reason for drivers to work. Without drivers, riders would have to be willing to wait forever. As long as a price pair admits an equilibrium with riders and drivers, I assume platforms reach this equilibrium. In practice, platforms would have to take action to avoid this chicken-and-egg dilemma. For example, platforms can initially guarantee drivers a minimum amount of earnings per hour, which may lead to negative profits in the short run.[24]

2.6 Profit- vs Welfare-maximizing prices

Now that we can compute the number of riders and drivers in equilibrium for a given pair of prices, we may proceed to determine the price levels that maximize profit and welfare. Let's start with welfare maximization. Total welfare equals riders' surplus minus the cost of vehicles, drivers and external congestion.[25] The pricing problem for a social planner in

[23]More sensible pricing implies higher q or lower p.

[24]Weyl (2010) proposes insulating tariffs (the price to one side depends on the number of agents on the other side) as a strategy for platforms to avoid coordination failure and implement the desired allocation.

[25]Since the number of riders and other road users are expressed in per-hour units, total welfare and profit are also given on a per-hour basis.

control of the platform can then be expressed as:

$$\max_{p,q} \quad \underbrace{N \int_0^\infty \int_{p+\beta(t(d)+w(x,d))}^\infty [V - \beta(t(d) + w(x,d))] f(\beta, V) \, \mathrm{d}V \, \mathrm{d}\beta}_{\text{Riders' surplus}} - \underbrace{c \cdot d}_{\substack{\text{Vehicle-driver} \\ \text{cost}}} - \underbrace{MEC \cdot d}_{\substack{\text{External} \\ \text{congestion cost}}} \tag{14}$$

where x and d are functions of p and q through the equilibration process analyzed in the previous section.

Appendix A shows that the welfare-maximizing price gap $(p - q)$ can be expressed as:

$$p - q = \underbrace{\bar{\beta} \cdot x \cdot \frac{\mathrm{d}(t+w)}{\mathrm{d}x}}_{\substack{\text{Marginal external} \\ \text{cost on ride hailers}}} + \underbrace{\frac{d}{x} \cdot MEC}_{\substack{\text{Marginal external cost} \\ \text{on other road users}}} \tag{15}$$

We can interpret Expression 15 as a Pigouvian tax. The welfare-maximizing price gap equals the sum of the marginal external cost an additional ride hailer imposes on her fellow ride hailers and on other road users. The marginal cost on ride hailers equals the product of their average value of time ($\bar{\beta}$), their quantity (x) and the marginal effect of an additional ride hailer on total travel time (in-vehicle plus wait, $\mathrm{d}(t+w)/\,\mathrm{d}x$).[26] This cost may turn out to be negative (a benefit) because additional ride hailers increase in-vehicle times through traffic congestion but also reduce wait times due to network effects (see Equation 9).[27] The marginal cost on other road users equals the value of the marginal increase in in-vehicle times caused by ride-hailing vehicles (MEC in Equation 10) multiplied by the ratio between vehicles and riders (d/x), which expresses the additional number of vehicles brought by one more rider.

Turning to profit maximization, profit per hour equals the product of the price gap and the number of trips per hour.[28] The pricing problem for a private platform has then a much

[26] $\mathrm{d}(t + w)/\,\mathrm{d}x$ is *not* a partial derivative. It takes into account the fact that an additional ride hailer causes a proportional increase in the number of ride-hailing vehicles (d).

[27] The optimal pricing strategy by a social planner may then imply a subsidy and negative revenue ($p - q < 0$), especially if the external cost on other road users is small. The potential need for subsidies in public transportation services due to network effects is well known in the transportation literature. See (Arnott, 1996) for taxis and (Parry & Small, 2009) for transit.

[28] I assume the costs of developing and maintaining the digital platform are either sunk or independent of the scale of use of the platform (i.e. fixed costs). Consequently, these costs do not affect the profit- or welfare-maximizing price levels.

simpler mathematical form:

$$\max_{p,q} \quad (p-q)\cdot x \tag{16}$$

where x is again a function of p and q through the equilibration process analyzed in section 2.5.

Appendix A shows that the profit-maximizing price gap can be expressed as:

$$p - q = \underbrace{\bar{\beta}_m \cdot x \cdot \frac{\mathrm{d}(t+w)}{\mathrm{d}x}}_{\substack{\text{Marginal external cost on ride}\\ \text{hailers valued according to the}\\ \text{value of time of marginal riders}}} + \underbrace{\frac{p}{\varepsilon}}_{\text{Markup}} \tag{17}$$

Expression 17 is again the sum of two terms. The first one is very similar to the first term of the welfare-maximizing price gap, but instead of the average value of time of riders ($\bar{\beta}$), it considers the average value of time of *marginal* riders ($\bar{\beta}_m$).[29] This type of distortion between welfare- and profit-maximizing pricing is not new. Its origins can be traced back to Spence (1975)'s analysis of quality provision by a monopolist. More recently, Weyl (2010) identifies the same distortion, which he names Spence distortion, for multi-sided platforms. It is then not surprising to find it for ride-hailing platform pricing. The second term (p/ε) can be interpreted as the usual markup imposed by a monopolist, which is proportional to the inverse of the elasticity of demand.[30]

A comparison of Expressions 15 and 17 reveals that the difference between the welfare- and profit-maximizing price gaps has three main sources. First, the indifference of a private platform towards the congestion effect on other road users. Second, the tendency of such platform to impose a markup. Third, the Spence distortion in the consideration of the external effect on other ride hailers. The first discrepancy tends to make the welfare-maximizing price gap larger, while the second one has the opposite effect. This contest between external effects and markup drives most of the comparison between price gaps. The Spence distortion can go either way, depending on the shape of the bivariate distribution $f(\beta, V)$. In particular, high correlation between reservation values a values of time pushes in favor of a larger welfare-maximizing price gap, because it tends to generate an average value of time of riders ($\bar{\beta}$) larger than that of marginal ones ($\bar{\beta}_m$).

It is interesting to note that a profit-maximizing platform internalizes (with a Spence distortion) external effects on ride hailers. Currently, ride-hailing vehicles account for only

[29]Marginal riders are those just indifferent between ride hailing and their outside option. They can be found along the ray of Figure 1.

[30]This measure of demand elasticity takes travel time as fixed. See Appendix A for more details.

a small portion of vehicle-miles traveled in most cities, so the platform internalizes only a small fraction of the external congestion effect on all road users. However, if ride hailing continues to grow, the platform will internalize a larger fraction of the externality, potentially weakening the motivation for a congestion charge.

3 Empirical estimates

This section presents the data and the empirical estimations carried out to calibrate the previous theoretical model to the morning peak period of Bogotá in 2019. I first describe the situation of ride hailing in Bogotá and then detail the two main empirical estimations (demand and marginal congestion).

3.1 Ride hailing in Bogotá

Bogotá is a highly dense city of about 7.5 million inhabitants distributed over an urban area of approximately 850km^2. Even though public transportation is the main mode of transportation in the city[31] and a license plate-based restriction takes 50% of the private car fleet out of circulation during peak hours (6-8:30am and 3-7:30pm), Bogotá usually tops worldwide traffic congestion rankings.[32]

Uber was the first ride-hailing platform in Bogotá, available since 2013. As in most countries, Uber has faced strong legal challenges in Colombia. The Ministry of Transportation declared ride-hailing services illegal because private vehicles are used to provide a public service, which goes against Colombian law, but the Ministry of Information and Communications Technologies refuses to block the apps based on net neutrality principles. Travelers cannot be penalized for using these services, but drivers can have their driver license suspended temporarily, their car withheld and face monetary penalties. In spite of these difficulties, other platforms such as Beat, Cabify and DiDi followed Uber and are currently available in Bogotá.

The 2019 Mobility Survey of Bogotá provides information about the size of ride hailing in the city for an average weekday (Secretaría Distrital de Movilidad, Bogotá D.C., 2019).[33]

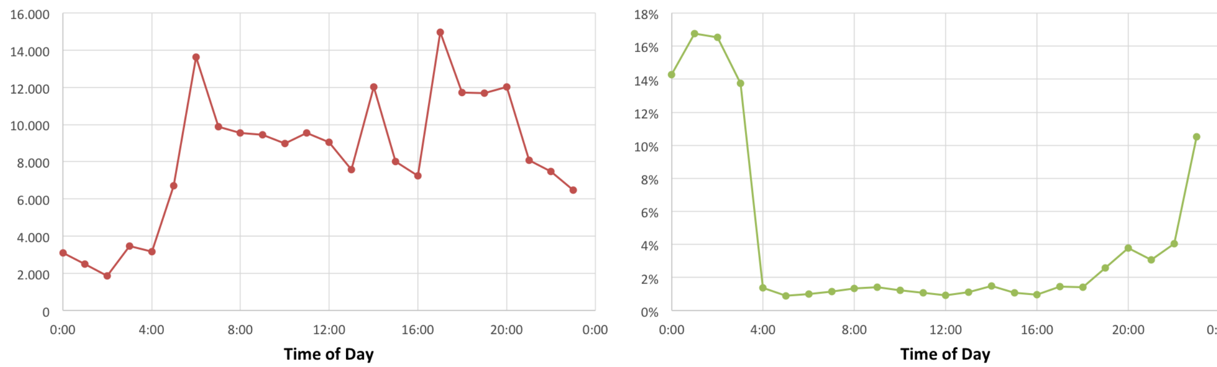
[31] Public transportation in Bogotá includes regular buses and a Bus Rapid Transit (TransMilenio) that operates on exclusive lanes.

[32] Two examples are the Global Traffic Scorecard (Inrix, 2019) and the global analysis based on Google Maps made by P. Akbar, Couture, Duranton, and Storeygard (2020). Both studies rank Bogotá as the most congested city in the world, although Chinese cities were not included.

[33] The survey recorded socio-demographic information from a sample of households, as well as detailed information about the trips made by household members the weekday before the data was collected. The sample consists of 21,828 households located in Bogotá and its surrounding municipalities, who were surveyed between February and August. The survey results include weights that make the sample representative of

Figure 3 reveals the number of ride-hailing trips made per hour throughout the day, as well as the percentage this number represents of all trips in the city.[34] Ride hailing peaks in the morning (6-7am) and in the evening (5-6pm) at about 14,000 trips per hour. It accounts for 1 to 2% of all trips most of the day, but its share increases to 15% after midnight (a pattern it shares with taxi trips).

Figure 3: Ride-hailing trips on an average weekday in Bogotá in 2019.



(a) Trips per hour. (b) Percentage of all trips.

Notes: Trips are classified according to their start times. For example, the number of trips at 6am corresponds to trips that started between 6:00am and 6:59am. The total number of trips considered to compute the percentages excludes walking trips shorter than 15 minutes.

To calibrate the theoretical model developed in Section 2, I consider the average size and characteristics of ride hailing during the morning peak period from 6:00 to 8:30am.[35] The average number of ride-hailing trips per hour during this period was 11,300. The average in-vehicle time of these trips was 37.7 minutes; their average wait time, 2.1 minutes; their average distance, 7.66 kilometers; and the average price paid by riders, COL$11,500. The survey also reveals that about 70% of ride-hailing trips during the morning peak used Uber.[36] Importantly, over 95% of ride hailers declared that they did not have a car available for their trip, which supports the model assumption of low substitution between ride hailing and

the population. Unless otherwise noted, all the statistics cited in this paper are weighted.

[34]To compute the total number of trips, I exclude walking trips shorter than 15 minutes.

[35]There are advantages and disadvantages to considering shorter (or longer) periods. The main advantage of shorter periods is that average conditions, such as demand and supply levels, are more likely to be representative of the entire period. The main disadvantage is that some of the model assumptions, such as the inability of riders to switch to a different time period, are less likely to hold. I chose a 2.5 hours period to strike a balance between these concerns. Additionally, the 6-8:30am period coincides with the effective period of the license plate-based restriction on private cars, which is important to maintain the model assumption of low substitution between ride hailing and private cars.

[36]Uber's dominance may have diminished in 2020 due to two events. First, DiDi joined the market in the second half of 2019. Second, Uber had to suspend its services for about a month at the beginning of 2020 due to allegations of illegal competition from taxi unions.

private cars. Finally, the survey shows that some areas of the city generate very few ride-hailing trips during the morning peak period. To account for this pattern, I consider a service area of 500km^2 (A in Equation 6).

Unfortunately, the survey does not reveal information on the side of drivers, such as the amount paid per trip by platforms or the number of vehicles available for service at different times. Fortunately, Azuara, Gonzalez, and Keller (2019) provide an extensive characterization of Uber drivers in Colombia and other Latin American countries. Based on Uber's administrative data, they report that drivers in Colombia generated an average hourly income of COL$14,075 using the platform in January-February 2019.[37] I use this amount as the reservation wage of drivers (c in Equation 3).[38] They also reveal that drivers use the platform in average 15 hours per week, which suggests that most drivers are available for service only a few hours per day. This statistic supports the model assumption that drivers can choose to work during the morning peak period without having to work on adjacent periods (i.e. driver supply is independent across time periods).

3.2 Demand

As explained in Section 2.1, a bivariate distribution $f(\beta, V)$ characterizes the population of potential ride hailers. This distribution represents the heterogeneity in values of time (β) and reservation values (V) across individuals. I assume the distribution has a bivariate normal form, which gives me five parameters to estimate: two means (μ_β, μ_V), two standard deviations (σ_β, σ_V) and one coefficient of correlation (ρ). Population size (N) will be adjusted to achieve, in a monopolistic scenario, the number of ride-hailing trips per hour estimated from the 2019 Mobility Survey.

I estimate these five parameters using data from stated-preference surveys carried out in Bogotá in December 2018 (Oviedo, Granada, & Perez-Jaramillo, 2020). In these surveys, individuals were asked to recall their most recent trip in the city during the morning peak period. They were then asked to choose between the mode of transportation actually used for the trip and a hypothetical ride-hailing alternative with a specific, and randomly assigned,

[37] The study reports an average hourly income of USD$10.5 adjusted by purchasing power parity (PPP). I use a PPP conversion rate of $1,340.5 COL/USD (OECD, 2019) to recover the amount in Colombian pesos. The average was based on the earnings (net of Uber's commission but inclusive of vehicle expenses such as fuel costs) of 1,136 drivers.

[38] The reservation wage of ride-hailing drivers during the morning peak period of Bogotá may differ from this amount due to three concerns. First, even though Bogotá is Uber's largest market in Colombia, the reporte may have included drivers in other Colombian cities. Second, reservation wages are likely to vary throughout the day. Chen et al. (2019) estimate, however, that reservation wage of Uber drivers in the U.S. during the morning peak period is close to the daily average. Finally, drivers may have different reservation wages for platforms other than Uber.

price (p), in-vehicle time (t) and wait time (w).[39] Let y be a binary variable that takes value 1 if the individual chose ride hailing, 0 otherwise. Each survey observation i can then be summarized by a vector (y_i, p_i, t_i, w_i). After data cleaning, I obtain 1,022 observations for estimation.

The surveys also collected socio-demographic information on each individual. I use these data to compute a weight for each observation (h_i) so that the sample resembles the characteristics of the population of travelers during the morning peak period according to the 2019 Mobility Survey. The characteristics considered to compute the weights include age, gender, socio-economic stratum, transportation mode, trip purpose and trip distance.[40]

The bivariate distribution $f(\beta, V)$ determines travelers' choices between ride hailing and their outside options given a price and in-vehicle travel time for an *average-distance* trip. I then adjust the price and and in-vehicle travel time (p_i, t_i) given to each individual in the survey according to their trip distance d_i and the average distance of ride-hailing trips obtained from the 2019 Mobility Survey ($\bar{d} = 7.66km$). For each survey observation, I compute $\tilde{p}_i = (\bar{d}/d_i)p_i$ and $\tilde{t}_i = (\bar{d}/d_i)t_i$.

I estimate the five parameters of the bivariate normal distribution through maximum likelihood. The log-likelihood of the entire sample given a set of parameters $\Theta = (\mu_\beta, \mu_V, \sigma_\beta, \sigma_V, \rho)$ is:

$$\mathcal{L}(\Theta) = \sum_{i=1}^{1,022} h_i \cdot \Bigg[y_i \cdot log \left(\int_0^\infty \int_{\tilde{p}_i + \beta(\tilde{t}_i + w_i)}^\infty f(\beta, V; \Theta) dV d\beta \right) + (1 - y_i) \cdot log \left(1 - \int_0^\infty \int_{\tilde{p}_i + \beta(\tilde{t}_i + w_i)}^\infty f(\beta, V; \Theta) dV d\beta \right) \Bigg] \quad (18)$$

where $f(\beta, V; \Theta)$ represents the bivariate normal probability density function.

Table 1 presents the point estimates and standard errors of the five parameters. The average value of time is COL\$9,100/hr, while the average reservation value is COL\$12,450. The results also reveal a positive correlation ($\hat{\rho} = 0.63$) between values of time and reservation values. It is reasonable to obtain a positive correlation because income level is probably a strong determinant of both values. Individual with higher incomes are expected to have larger values of time and reservation values. Figure 4 graphs the bivariate distribution based on the point estimates.

[39] In case the individual actually made the trip by ride hailing, she was asked if she would continue to do so under hypothetical characteristics for ride hailing p, t and w.

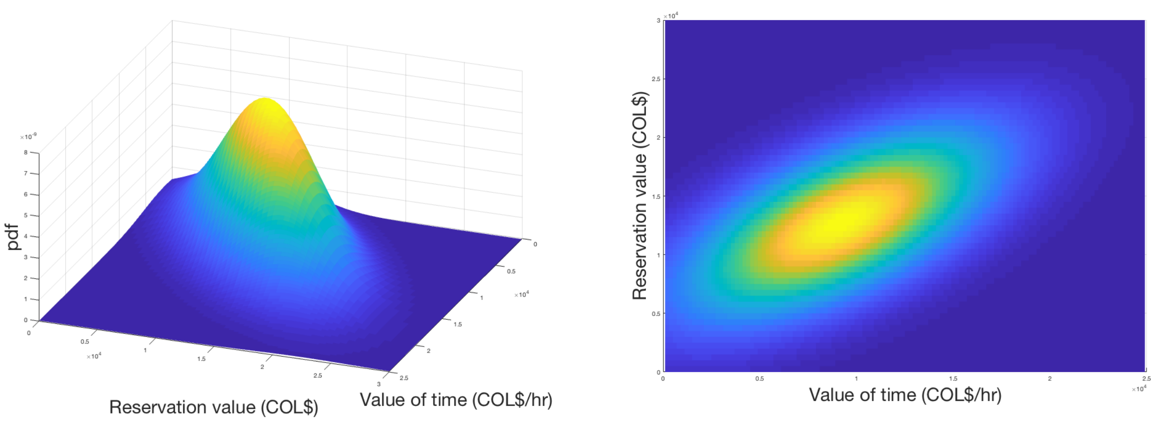
[40] I applied an iterative raking procedure to compute the weights.

Table 1: Parameter estimates for the bivariate normal distribution $f(\beta, V)$.

	μ_β	σ_β	μ_V	σ_V	ρ
Estimate	$9,100/hr	$5,220/hr	$12,450	$5,270	0.63
Std. Err.	$1,160/hr	$1,640/hr	$660	$540	0.15

Notes: All monetary values are in Colombian pesos. Parameters were estimated through maximum likelihood. Standard errors were computed using the Cramer-Rao bound.

Figure 4: Estimated bivariate normal distribution $f(\beta, V)$.



(a) 3D view. (b) View from above.

3.3 Traffic congestion

The main objective of the empirical work presented in this section is to identify the marginal effect of additional vehicles on the streets (traffic volume) on average in-vehicle travel times during the morning peak period in Bogotá. In-vehicle travel times for a given time period result as an equilibrium outcome of the interaction between travel demand (measured in number of vehicles), which declines as travel times increase, and the capacity of the road network, which dictates how travel times rise as traffic volume grows. To identify the second effect (commonly referred to as the supply side of transportation), we can then use exogenous changes in travel demand, which occur naturally throughout the day as people prefer to (or must) travel at specific times of day.[41]

To illustrate the relationship between travel speed and traffic volume throughout the day, Figure 5 displays estimates of both for every 5-min interval of the day based on data from the 2019 Mobility Survey. The figure shows that average speeds oscillate mostly between 20 and 30 km/hr between midnight and 4am, when traffic volume is below 10,000 vehicles, and diminish to about 12 km/hr at peak times, when the number of vehicles rises above 100,000. Traffic volume varies between 70,000 and 100,000 vehicles during most of the day, while average speed stays around 15 km/hr.

Besides traffic volume, several trip or traveler characteristics may differ across trips taken at different times of day. If these characteristics affect travel speed, they may lead to biases in simple regressions between speed and traffic volume. For example, most trips made during peak periods have mandatory purposes (work or study), and these purposes may encourage travelers to drive faster. To account for these potential effects, I run regressions between speed and traffic volume at the trip level, while controlling for diverse trip and traveler characteristics. The base specification for the regressions is:

$$\frac{1}{speed_i} = \alpha + \beta_v \cdot Veh_i + \vec{\beta}_C \cdot Controls_i + \varepsilon_i \tag{19}$$

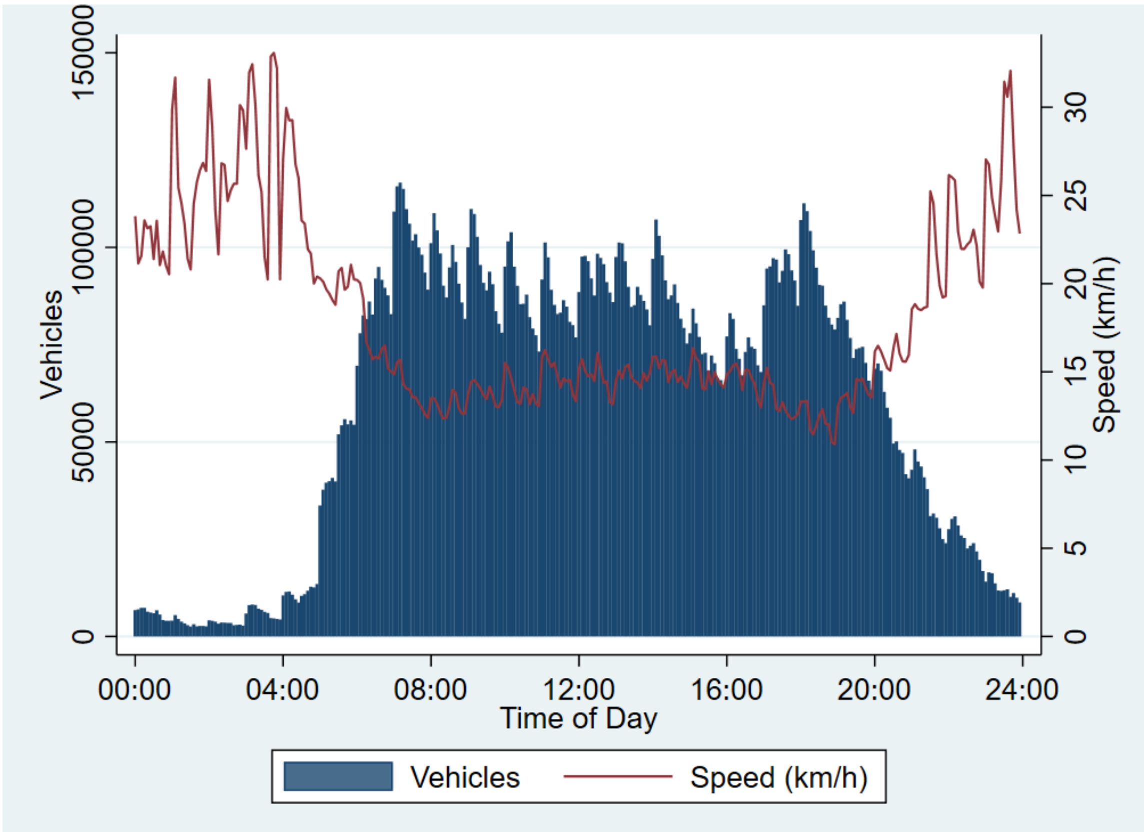
where $speed_i$ denotes the speed of trip i, Veh_i represents average traffic volume for the duration of trip i,[42] $Controls_i$ include a set of control variables related to the trip (distance, purpose and mode) and the traveler (age, gender and socio-economic stratum), and ε_i is an error term. The dependent variable for the regressions is the inverse of speed, which is proportional to in-vehicle travel time.

The relationship between traffic volume and travel time is usually found to be nonlinear,

[41] The empirical approach employed in this section was originally proposed and applied to Bogotá by P. A. Akbar and Duranton (2017).

[42] I average the number of vehicles on the streets in all 5-min intervals at least partially covered by the trip.

Figure 5: Traffic volume and average speed in Bogotá (average weekday).



Notes: The number of vehicles includes private cars, taxis and ride-hailing vehicles. The 2019 Mobility Survey differentiates private car trips as driver and as passenger. To compute the number of private cars, I consider only car trips as driver. For taxis and ride-hailing vehicles, I assume one trip corresponds to one vehicle. This assumption may overestimate the number of vehicles to the extent that passengers share rides, but it also may underestimate it to the extent that vehicles deadhead (travel without a passenger). I include trips in a 5-min interval as long they cover any portion of the interval. For example, a trip starting at 8:37am and ending at 9:11am is included in the eight 5-min intervals between 8:35am and 9:15am. I subtract any recorded wait and walk times at origin or destination to consider only in-vehicle travel time. Speed observations come from the same modes. The survey does not report the distance covered by each trip. I approximate these distances by querying Google Maps and obtaining the distance of the recommended route between origin and destination under average traffic conditions. The speed of a 5-min interval corresponds to the average speed of all trips that cover any portion of the interval. Average speeds fluctuate more between midnight and 4am mainly due to less observations for this period.

with travel times increasing more rapidly as traffic volume grows (Small & Verhoef, 2007). To approximate this nonlinear relationship in a flexible manner, I introduce a piecewise linear specification for Veh_i in Equation 19. In this specification, the marginal effect of additional vehicles on the inverse of speed (or travel time) may vary for different ranges of traffic volume.

Table 2 presents regression results for four different versions of Equation 19. Column (4) contains the preferred specification, which includes controls and introduces the number of vehicles in a piecewise linear form. The coefficients show that additional vehicles do not affect speed when traffic volume is below 20,000 vehicles, while they have the greatest impact when it rises above 80,000 vehicles. Since I calibrate the theoretical model to the morning peak period, when traffic volumes are above 80,000, the coefficient of interest corresponds to the one for this last range of traffic volume. Its point estimate is 6.98 x 10^{-7} hr/km. This magnitude implies that 10,000 additional vehicles on the streets increase in-vehicle travel time by about 4.2 minutes for a 10 km trip.

Table 2: Regression results - Inverse of speed on traffic volume.

		(1)	(2)	(3)	(4)
Vehicles		5.22^{***} (0.25)		4.29^{***} (0.25)	
Vehicles	<20,000		1.72 (3.80)		0.08 (3.58)
	20,000-40,000		5.50^{*} (2.82)		4.18 (2.66)
	40,000-60,000		3.90^{*} (2.29)		3.74^{*} (2.15)
	60,000-80,000		6.64^{***} (1.67)		4.04^{**} (1.58)
	>80,000		5.64^{***} (1.30)		6.98^{***} (1.24)
Controls		No	No	Yes	Yes

Notes: Each regression estimates Equation 19 at the trip level, with the inverse of speed as the dependent variable. All regressions include a constant (α). Columns (1) and (3) include the number of vehicles (Veh_i) as a single variable. Columns (2) and (4) include it in a piecewise linear specification with five ranges. The units of all coefficients are 10^{-7} hr/km. Controls include trip distance and dummy variables that identify the purpose of the trip (mandatory or discretionary), its mode (car, taxi or ride hailing), and the age (<30, 30-50 or >50), gender and socio-economic stratum of the traveler. Standard errors in parenthesis. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

As introduced in Equation 10, the marginal external cost (MEC) each ride-hailing vehicle imposes on other road users through traffic congestion can be approximated as the product of the number of users, their average value of time and the average in-vehicle travel time increment caused by an additional vehicle. I consider two groups of other road users: private cars and taxis.[43] I estimate the number of car and taxi users per hour during the morning

[43] I ignore the potential impact on public transportation users. TransMilenio uses exclusive bus lanes,

peak period based on the 2019 Mobility Survey. For both groups, I use the average value of time estimated in Section 3.2 for the population of potential ride hailers (μ_β in Table 1). Finally, the average in-vehicle travel time increment caused by an additional vehicle for each group equals the product of their average trip distance (9.3 and 7.5 km for car and taxi trips respectively) and the coefficient obtained from the previous regressions (6.98 x 10^{-7} hr/km).

Similarly, I approximate the in-vehicle travel time function for ride hailing trips (t(d) in Expression 1) as the sum of a base time and the increment caused by ride-hailing vehicles through traffic congestion.[44] Finally, I assume ride-hailing vehicles travel at a constant speed of 20 km/hr (not affected by congestion) when en route to pick up a passenger (v in Equation 5).[45]

4 Results

Table 3 reveals the main results of this paper. It presents the optimal pricing decisions of the ride-hailing platform in three scenarios, as well as the outcomes generated by these decisions. In the first scenario, a profit-maximizing firm manages the platform. In the second, it is managed by a social planner, who attempts to maximize overall welfare and so internalizes external congestion effects on other road users. Finally, a private firm takes back control of the platform in the third scenario, but a regulator imposes a tax on the price charged to riders.

The first scenario resembles the market structure of ride hailing in Bogotá in 2019. As mentioned in Section 3.1, even though Uber was not the only platform available in Bogotá, it controlled the majority of the market. The platform charges riders a price per trip of COL$12,060, while it pays drivers COL$9,640. These values represent prices for an average-distance trip, while the price for specific trips is adjusted in proportion to distance. Since the average distance of a ride-hailing trip during the morning peak period was 7.66 km, the platform's pricing strategy can also be interpreted as charging riders COL$1,575 per km and paying drivers COL$1,260 per km. According to the 2019 Mobility Survey, the average price paid by ride hailers during the morning peak period was about COL$11,500. The

while the speed of regular buses depends mainly on the number of stops required to pick up and drop off passengers.

[44]The base time represents the expected in-vehicle travel time for a ride-hailing trip of average distance. I adjust this base time so that in-vehicle time in the monopolistic scenario matches the average observed in the 2019 Mobility Survey. The time increment caused by ride-hailing vehicles equals again the average trip distance (7.66km for ride hailing) and the coefficient obtained from the regressions.

[45]This assumption reflects the notion that vehicles use mainly local uncongested roads when en route to pick up a passenger.

Table 3: Numerical results for the three main scenarios.

		Monopoly	Social planner	Tax=$1,390
Prices	Riders (p)	$12,060	$13,060	$11,740
	Drivers (q)	$9,640	$9,620	$9,670
Platf. Commission		20.1%	26.4%	17.6%
Quantities	Trips per hour (x)	11,130	7,000	6,980
	Vehicles (d)	7,620	4,780	4,800
Traffic volume increase		8.5%	5.3%	5.3%
In-vehicle time increase		3.7%	2.3%	2.3%
Vehicle utilization		91.3%	89.3%	89.9%
Wait time (min)		2.8	3.7	3.2
Profit (millions/hr)		$26.9	-	$14.4
Public Revenue (millions/hr)		-	$24.1	$9.7
Welfare (millions/hr)		$9.6	$12.1	$12.0

Notes: All monetary figures are in Colombian pesos. Prices apply to an average-distance trip. The price faced by riders in the last column is COL$13,130 (=$11,740+$1,390).

monopolistic scenario slightly overestimates this value.[46]

The monopolist platform imposes a price gap or platform commission of COL$2,420 or 20.1% (measured as a percentage of the price charged to riders).[47] This gap can be interpreted as the size of the markup imposed by the platform. As a result of these prices, the number of riders or trips taken per hour is 11,130, while the number of drivers or vehicles available for service is 7,620.[48] Ride-hailing vehicles cause an increase in traffic volume of 8.5%, which raises in-vehicle travel times of all road users by 3.7% in average.[49] The utilization rate of ride-hailing vehicles, measured as the percentage of time vehicles have a passenger on board,

[46]This difference may be due to the monopolistic market structure imposed on the model, in comparison to a slightly competitive market in reality. But it may also reflect inaccuracies in theoretical assumptions, estimated parameters or Uber's attention to other objectives besides short-run profits (such as market growth).

[47]The commission charged by Uber varies for each trip, but it is thought to be around 20 to 25%. Other platforms, such as DiDi and Cabify, claim to charge lower commissions around 10 to 15%.

[48]I adjusted the size of the population of potential travelers (N) so that the trips per hour in this scenario reflect the average observed in the 2019 Mobility Survey. The final population size was $N = 135,000$.

[49]Balding, Whinery, Leshner, and Womeldorff (2019) reveal that Uber and Lyft account for 2 to 13% of vehicle-miles traveled (VMT) at the core counties of 6 major U.S. cities. The 8.5% increase in traffic volume I find for Bogotá, although indicative, is not entirely comparable to these percentages because I do not include VMT by buses and freight vehicles as part of the total.

is 91.3%.[50] The average wait time experienced by riders is 2.8 minutes, which is slightly higher than the 2.1 minutes observed from the 2019 Mobility Survey. Finally, the platform gains profits of COL$26.9 millions per hour during the morning peak, while the availability of ride hailing increases overall welfare by COL$9.6 millions per hour. Overall welfare includes net welfare gains for ride hailers minus the external congestion cost imposed on other road users (Equation 14). In this case, the net benefit created for ride hailers outweighs external congestion costs.

When a social planner takes control of the platform, her main action is to raise the price charged to riders to COL$13,060, while maintaining the price paid to drivers at about the same level. The price gap then rises to COL$3,440, which constitutes a platform commission of 26.4%. This gap represents the marginal external cost an additional ride hailer imposes on other road users.[51] The price increase applied by the social planner has the expected effects. The number of riders and drivers decreases by about 37%. The effect on traffic congestion declines, while average wait times rise. Ride hailing now increases overall welfare by more than COL$12 millions per hour, a 26% increase from the monopolistic scenario. Finally, the platform obtains lower profits (10% reduction), which now constitute public revenue.

The results then suggest that the size of the markup was less than the marginal external cost for Bogotá's ride-hailing market during the morning peak period in 2019.

A social planner does not have to take control of the platform to realize all the potential welfare gains from ride hailing. In the third scenario, I compute the optimal tax a regulator should impose on the price charged to riders by a profit-maximizing platform in order to maximize overall welfare. The size of the optimal tax or congestion charge is COL$1,390. Again, this value corresponds to the optimal charge for an average-distance trip, while the charge for specific trips should be adjusted in proportion to distance. In other words, the optimal charge should be interpreted as about COL$180 per kilometer. Alternatively, the optimal charge can be applied as a 12% tax on the price charged by the platform to riders.

The size of the optimal charge is larger than the difference between the optimal price gaps in the first two scenarios (COL$1,020). There is a good reason for this discrepancy. The monopolist platform reduces the price it collects from riders as a response to the tax, causing an incomplete pass-through of the tax to riders. In the unregulated scenario, the platform charges riders COL$12,060 per trip, but when the the regulator imposes the tax

[50] Cramer and Krueger (2016) and Balding et al. (2019) report vehicle utilization rates for Uber and Lyft in U.S. cities between 50 and 70%. The utilization rate I find for Bogotá is relatively high in comparison. Two important differences may explain this disparity. First, Bogotá is significantly denser than most U.S. cities, which probably leads to better matching (in terms of distance) between riders and vehicles. Second, I focus on the peak period, while the statistics reported for U.S. cities include off-peak periods and weekends.

[51] The marginal cost on other ride hailers (first term of Expression 15 turns out to be mostly negligible. The effect on traffic congestion offsets the effect on wait time.

the platform decides to collect only COL$11,740 per trip. As a result, the price faced by riders increases from COL$12,060 to COL$13,130 (=$11,740+$1,390) due to the tax. These results imply a pass-through of 0.77.

In this regulated scenario, ride-hailing increases overall welfare by COL$12 millions per hour, which is very close to the maximum possible (COL$12.1 millions per hour, achieved when a social planner controls the platform). For the regulator to achieve the maximum welfare increase possible, she would have to regulate on both sides of the market. However, there is little incentive for her to additionally regulate the side of drivers, because the best possible tax on the side of riders achieves over 95% of the welfare increase available from the unregulated scenario.

Not surprisingly, the tax reduces the profit of the platform. Private profit decreases by about 46% (from COL$26.9 to COL$14.4 millions per hour). Finally, the tax generates an important public revenue of almost COL$10 millions per hour during the morning peak period.

5 Conclusions

By electronically matching riders and drivers, digital platforms raised the efficiency of ride-hailing services and consequently increased their use by urban travelers. Unfortunately, the rapid growth of these services threatens to exacerbate transportation-related externalities in cities around the world, most importantly traffic congestion. Economists have long argued for congestion charges as a tool to mitigate congestion externalities, but their application to private cars has materialized in only a few cities around the world. However, congestion charges may prove easier to implement for modern ride-hailing services, because platforms already have the technology in place to identify and charge individual trips.

Ride-hailing markets are likely to gravitate towards high levels of concentration in very few digital platforms, mainly because of positive network effects in wait times. As platforms exert market power to dictate prices, we will move into the terrain of environmental regulation under market power. In this terrain, it is not optimal for private firms to completely internalize external damages.

In this paper, I present the first comparison between market power and congestion externalities for ride-hailing markets. For the morning peak period of Bogotá in 2019, I find that the marginal external cost of congestion is larger than the markup imposed by a monopolist platform. A congestion charge on ride hailing is then justified. However, the size of this charge is only 40% of the marginal external cost of congestion. These conclusions can change significantly for other cities, and even for Bogotá at different time periods, as the factors that affect congestion externalities and market power vary. Some of the most important factors include the elasticity of demand for ride-haling services, its level of substitution with private cars and the effect of additional vehicles on travel speeds.

References

Akbar, P., Couture, V., Duranton, G., & Storeygard, A. (2020, October). *Mobility and congestion in world cities: Evidence from google maps.* (Presented at the NBER Conference Economics of Transportation in the 21st Century)

Akbar, P. A., & Duranton, G. (2017). *Measuring the cost of congestion in a highly congested city: Bogotá* (Working paper No. 2017/04). CAF - Development Bank of Latin America.

Arnott, R. (1996). Taxi travel should be subsidized. *Journal of Urban Economics*, *40*, 316-333.

Azuara, O., Gonzalez, S., & Keller, L. (2019). *Who drives on ride-hailing platforms in latin america?* (Technical Note No. IDB-TN-1779). Inter-American Development Bank.

Balding, M., Whinery, T., Leshner, E., & Womeldorff, E. (2019). *Estimated TNC share of VMT in six US metropolitan regions* (Tech. Rep.). Fehr and Peers.

Bian, B. (2018, April). *Search frictions, netwrok effects and spatial competition: Taxis versus uber.*

Brueckner, J. K. (2002). Airport congestion when carriers have market power. *American Economic Review*, *92*(5), 1357-1375.

Buchanan, J. M. (1969). External diseconomies, corrective taxes, and market structure. *American Economic Review*, *59*(1), 174-177.

Buchholz, N. (2020, February). *Spatial equilibrium, search frictios and dynamic efficiency in the taxi industry.*

Castillo, J. C. (2019, December). *Who benefits from surge pricing?*

Castillo, J. C., Knoepfle, D., & Weyl, E. G. (2018). *Surge pricing solves the wild goose chase.*

Chen, M. K., Chevalier, J. A., Rossi, P. E., & Oehlsen, E. (2019). The value of flexible work: Evidence from uber driers. *Journal of Political Economy*, *127*(6).

Cohen, P., Hahn, R., Hall, J., Levitt, S., & Metcalfe, R. (2016). *Using big data to estimate consumer surplus: The case of uber.* (NBER Working Paper Series)

Cramer, J., & Krueger, A. B. (2016). Disruptive change in the taxi business: The case of uber. *American Economic Review: Papers and Proceedings*, *106*(5), 177-182.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science Advances*, *5*(5).

Fowlie, M., Reguant, M., & Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, *104*(1).

Frechette, G. R., Lizzeri, A., & Salz, T. (2019). Frictions in a competitive, regulated market: evidence from taxis. *American Economic Review*, *109*(8), 2954-2992.

Hall, J. D., Palsson, C., & Price, J. (2018). Is uber a substitute or complement for public transit? *Journal of Public Economics*, *108*, 36-50.

Hall, J. V., Horton, J. J., & Knoepfle, D. T. (2020, June). *Ride-sharing markets re-equilibrate.* (Unpublished)

Inrix. (2019). *Inrix global traffic scorecard.* Retrieved from https://inrix.com/scorecard/

Lam, C. T., & Liu, M. (2017). *Demand and consumer surplus in the on-demand economy: the case of ride sharing* (Woring paper). MIT Initiative on the Digital Economy.

Li, S., Tavafoghi, H., Poolla, K., & Varaiya, P. (2019). Regulating TNCs: Should Uber and Lyft set their own rules? *Transportation Research Part B*, *129*, 193-225.

Mills, D. E. (1981). Ownership arrangements and congestion-prone facilities. *American Economic Review*, *71*(3), 493-502.

New York City Taxi and Limousine Commission. (2019). *New york state's congestion surcharge.* Retrieved from https://www1.nyc.gov/site/tlc/about/congestion-surcharge.page

Oviedo, D., Granada, I., & Perez-Jaramillo, D. (2020). Ridesourcing and travel demand: Potential effects of transportation network companies in bogotá. *Sustainability*, *12*(5).

Parry, I. W. H., & Small, K. A. (2005). Does Britain or the United States have the right gasoline tax? *American Economic Review*, *95*(4), 1276-1289.

Parry, I. W. H., & Small, K. A. (2009). Should urban transit subsidies be reduced? *American Economic Review*, *99*(3), 700-724.

San Francisco County Transportation Authority. (2017). *TNCs Today* (Tech. Rep.). Retrieved from https://www.sfcta.org/emerging-mobility/tncs-today

Secretaría Distrital de Movilidad, Bogotá D.C. (2019). *Encuesta de movilidad 2019.* Retrieved from https://www.simur.gov.co/portal-simur/datos-del-sector/encuestas-de-movilidad/

Shapiro, M. H. (2018, May). *Density of demand and the benefit of uber.* (Job market paper)

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, *1*, 2-14.

Small, K. A., & Verhoef, E. T. (2007). *The economics of urban transportation.* Routledge.

Spence, A. M. (1975). Monopoly, quality, and regulation. *Bell Journal of Economics*, *6*(2), 417-429.

Statista. (2020). *Market share of the leading ride-hailing companies in the united states.* Retrieved from https://www.statista.com/statistics/910704/market-share-of-rideshare-companies-united-states/

Tirachini, A. (2020). Ride-hailing, travel behavior and sustainable mobility: an international review. *Transportation*, *47*(4), 2011-2047.

Verhoef, E. T., & Small, K. A. (2004). Product differentiation on roads: Constrained congestion pricing with heterogeneous users. *Jorunal of Transport Economics and Policy*, *38*(1), 127-156.

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

Yanocha, D., & Mason, J. (2019). *Ride Fair: A policy framework for managing Transportation Network Companies* (Tech. Rep.). Institute for Transportation and Development Policy.

Appendix

A Profit- vs Welfare-maximizing prices

This appendix derives the expressions for the welfare- and profit-maximizing price gaps analyzed in section 2.6 (Equations 15 and 17). It also presents the additional first-order condition that is needed in each problem to determine both prices.

A.1 Welfare maximization

The welfare maximization problem can be stated mathematically as (section 2.6):

$$\max_{p,q} \quad N\int_0^\infty \int_{p+\beta(t(d)+w(x,d))}^\infty [V-\beta(t(d)+w(x,d))]f(\beta,V)\,\mathrm{d}V\,\mathrm{d}\beta - c\cdot d - MEC\cdot d \tag{20}$$

where x and d are implicit functions of p and q through the following two equations (sections 2.1 and 2.2):

$$N\int_0^\infty \int_{p+\beta(t(d)+w(x,d))}^\infty f(\beta,V)\,\mathrm{d}V\,\mathrm{d}\beta = x \tag{21}$$

$$d = \frac{q}{c}x \tag{22}$$

We can take d out of the problem using Equation 22. The maximization problem then turns to:

$$\max_{p,q} \quad N\int_0^\infty \int_{p+\beta\left(t\left(\frac{q}{c}x\right)+w\left(x,\frac{q}{c}x\right)\right)}^\infty \left[V - \beta\left(t\left(\frac{q}{c}x\right) + w\left(x, \frac{q}{c}x\right)\right)\right] f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta - q\cdot x - MEC\cdot\frac{q}{c}x \tag{23}$$

while x becomes an implicit function of p and q through the following equation:

$$N\int_0^\infty \int_{p+\beta\left(t\left(\frac{q}{c}x\right)+w\left(x,\frac{q}{c}x\right)\right)}^\infty f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta = x \tag{24}$$

The first order condition of problem 23 with respect to p is:

$$\begin{aligned} &-\frac{\mathrm{d}x}{\mathrm{d}p}\left(t_d\frac{q}{c} + w_x + x_d\frac{q}{c}\right) N\int_0^\infty \int_{p+\beta(t+w)}^\infty \beta f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta \\ &- N\int_0^\infty p\left[1 + \beta\frac{\mathrm{d}x}{\mathrm{d}p}\left(t_d\frac{q}{c} + w_x + x_d\frac{q}{c}\right)\right] f(\beta, p+\beta(t+w))\,\mathrm{d}\beta - q\frac{\mathrm{d}x}{\mathrm{d}p} - MEC\frac{q}{c}\frac{\mathrm{d}x}{\mathrm{d}p} = 0 \end{aligned} \tag{25}$$

where t_d, w_x and w_d denote the partial derivatives of the in-vehicle and wait time functions with respect to the number of riders and vehicles. From equation 24 and the implicit function theorem, the derivative of the number of riders with respect to price is:

$$\frac{\mathrm{d}x}{\mathrm{d}p} = \frac{-N\int_0^\infty f(\beta, p+\beta(t+w))\,\mathrm{d}\beta}{1 + \left(t_d\frac{q}{c} + w_x + x_d\frac{q}{c}\right) N\int_0^\infty \beta f(\beta, p+\beta(t+w))\,\mathrm{d}\beta} \tag{26}$$

Plugging Expression 26 into Equation 25 and noting that the average value of time of riders can be expressed as:

$$\bar{\beta} = \frac{\int_0^\infty \int_{p+\beta(t+w)}^\infty \beta f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta}{\int_0^\infty \int_{p+\beta(t+w)}^\infty f(\beta, V)\,\mathrm{d}V\,\mathrm{d}\beta} \tag{27}$$

the first-order condition can be written as:

$$p = \bar{\beta}x\left(t_d\frac{q}{c} + w_x + x_d\frac{q}{c}\right) + q + \frac{q}{c}MEC \tag{28}$$

Expressing the term in parentheses as $\mathrm{d}(t+w)/\mathrm{d}x$, the (not partial) derivative of travel time with respect to the number of drivers, noting that $q/c = d/x$ from Equation 22 and moving q to the left-hand side, one obtains Equation 15, which is the one analyzed in section 2.6.

The first-order condition for q, after a similar but slightly more cumbersome process, can be expressed as:

$$c + \bar{\beta} \cdot x \cdot t_d + MEC = -\bar{\beta} \cdot x \cdot w_d \tag{29}$$

We can interpret this condition as follows. The left-hand side reveals the cost of having one more ride-hailing vehicle, which includes the direct cost of vehicle expenses and driver labor (c), the external congestion cost on ride hailers ($\bar{\beta} x t_d$) and the external congestion cost on other road users (MEC). The right-hand side measures the benefit of that additional vehicle, which is the reduction in wait time for ride hailers ($-\bar{\beta} x w_d$). At an optimal solution, these marginal costs and benefit must be equal.

A.2 Profit maximization

The profit maximization problem is relatively simpler:

$$\max_{p,q} \quad (p - q) \cdot x \tag{30}$$

while x is again an implicit function of p and q through Equation 24.

The first order condition of problem 30 with respect to p is:

$$x + (p - q) \cdot \frac{\mathrm{d}x}{\mathrm{d}p} = 0 \tag{31}$$

Plugging Expression 26 into Equation 31 and noting that the average value of time of *marginal* riders can be expressed as:

$$\bar{\beta}_m = \frac{\int_0^\infty \beta f(\beta, p + \beta(t + w))\,\mathrm{d}\beta}{\int_0^\infty f(\beta, p + \beta(t + w))\,\mathrm{d}\beta} \tag{32}$$

the first-order condition can be written as:

$$p = \bar{\beta}_m x \left(t_d \frac{q}{c} + w_x + x_d \frac{q}{c} \right) + \frac{x}{N \int_0^\infty f(\beta, p + \beta(t + w))\,\mathrm{d}\beta} + q \tag{33}$$

Expressing the term in parentheses as $\mathrm{d}(t+w)/\mathrm{d}x$ (as in the welfare-maximization problem), we move to:

$$p = \bar{\beta}_m x \frac{\mathrm{d}(t + w)}{\mathrm{d}x} + \frac{x}{N \int_0^\infty f(\beta, p + \beta(t + w))\,\mathrm{d}\beta} + q \tag{34}$$

From Equation 24 (or more directly from Expression 26), the derivative of the number of riders with respect to price when travel time $(t + w)$ is taken as fixed is:

$$\left.\frac{\mathrm{d}x}{\mathrm{d}p}\right|_{t+w} = -N \int_0^{\infty} f(\beta, p + \beta(t + w))\,\mathrm{d}\beta \tag{35}$$

The absolute value of the elasticity of the number of riders with respect to price, again holding travel time fixed, can then be expressed as:

$$\varepsilon = \frac{p}{x} \cdot N \int_0^{\infty} f(\beta, p + \beta(t + w))\,\mathrm{d}\beta \tag{36}$$

Using this last expression to simplify the second term on the right-hand side of Equation 34, and moving q to the left-hand side, one obtains Equation 17, which is the one analyzed in section 2.6.

After a similar process, the first-order condition for q can be expressed as:

$$c + \bar{\beta}_m \cdot x \cdot t_d = -\bar{\beta}_m \cdot x \cdot w_d \tag{37}$$

Comparing this condition to the equivalent one for welfare maximization (Equation 29), we note that the profit-maximizing platform does not take into account the external congestion cost on other road users. Additionally, it values the external congestion cost on ride hailers using the average value of time of *marginal* riders (Spence distortion). As mentioned on section 2.2, there is no markdown distortion on this side of the market (the side of drivers) due to the uniform-reservation-wage assumption.