The price is right!

Information and dynamics in online marketplaces.

Rossi Abi-Rafeh* Emil Palikot†

November 19, 2019

Preliminary and incomplete

Click here for latest version

Abstract

A fundamental characteristic of many major online marketplaces is a fast turnover of sellers. This article studies the main factors influencing sellers' pricing and entry and exit decisions of sellers in a market with a reputation system. We provide a model of dynamic oligopoly with heterogeneity in costs and reputation levels across sellers to show that: entrants are generally more likely to leave the platform then incumbents, sellers who have a higher chance of leaving set on average higher prices and that conditioned on staying sellers increase their prices. We first simulate the model and show when our predictions are satisfied. Finally, we use a dataset on sellers on a popular ridesharing platform, to show that general patterns are consistent with our theory, and to calibrate our model in order to generate counterfactuals.

*Toulouse School of Economics. Email: rossi.abi-rafeh@tse-fr.com . This draft was written while Rossi was visiting Harvard's economics departement with funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228.

†Toulouse School of Economics. Email: emil.palikot@gmail.com. Emil Palikot acknowledges support from the European Research Council under the Grant Agreement no. 340903

1 Introduction

56% of people offering their capital to rent on sharing economy platforms (airbnb, etc...) quit within their first year, 52% of those offering their labour services on a sharing platform do as well.[1] Understanding what drives the exit decisions of sellers is crucial to comprehend market dynamics, and to predict under which conditions tools used by online platforms to promot certain groups of sellers are effective. A nascent online marketplace typically focuses on growing a seller base in hope of triggering network effects and attracting more buyers; once such a market becomes established, the market-maker has to carefully balance the advantages given to entrants (in form of for example increased prominence) against benefits of staying on the platform and becoming an incumbent.

This paper provides a model of sellers' behavior that, shows that unobserved, by econometricians, sellers' heterogeneity in form of opportunity costs, is crucial to understand entry, exit as well as pricing decisions. The presence of a reputation system introduces a static-dynamic trade-off in sellers behavior: a seller would enter the market if she can set a high enough price (i.e. if the market is profitable for her), however lower prices in the early listing can increase her reputation quickly and increase her future profits. might decide to enter the market only because it will benefit her in future. Several features of market design influence the pricing and entry decision of the seller. If, for example, platform promotes entrants, their incentive to invest in their future is lowered, but short-term profits higher.

We construct a dynamic model of oligopolistic competition à la Ericson and Pakes (1995). Sellers compete in prices and make decisions to exit the market. Seller's reputations develop following a probabilistic process. We introduce two dimensions of heterogeneity in costs: marginal costs which are observed by all sellers and constant over time, and opportunity costs, which differ from period to period. Numerical simulations of our preliminary model show that, sellers with a high marginal cost set higher prices; however, they leave the platform the moment they receive a high draw of the opportunity cost. Sellers with low marginal costs remain on the platform and start building a reputation, which allows them to start gradually increasing the prices. Relatively high heterogeneity in marginal costs is key to generate such a price dispersion.

We test predictions of our model using data from Blablacar, a major online ride-sharing platform. Where we show that entrant sellers set higher prices than incumbents, despite the accumulated rep-

[1] JP Morgan Chase Institute

utation. After controlling for a rich set of driver and listing-specific controls we find that the average effect of an extra review on the price is negative and significant. Second, in order to provide a reduced-form test of our model we account for unobserved heterogeneity in seller-specific characteristics. In a panel data study, we recover the expected price dynamic: sellers generally start at a lower price and increase it gradually. Both the first-difference (FD) and within (FE) estimators show a positive and significant impact of reputation on prices: going from 0 to 10 reviews now *increases* the average price by 70 cents. The Hausman test rejects the uncorrelatedness of the driver-specific unobservable effects with the number of reviews and the other observables. Thus, we show that the observed behavior is consistent with our theoretical model.

The paper proceeds as follows. In section 2 we discuss related literature, section 3 introduces our model and shows simulations of it. Section 4 discusses the empirical test of our predictions. Section 5 concludes.

2 Literature review

This project tackles the problem of optimal dynamic pricing by firms competing on a platform, in the presence of incomplete information about their quality. Hence, it falls into several strands of economic literature. First, pricing by long-lived economic agents has been a subject of study by both theoretical and empirical literature. Maskin and Tirole (1988a,b) have introduced the equilibrium concept on which most of the literature on dynamic oligopoly is based. These papers describe features of a Markov Perfect Equilibrium (MPE) and derive dynamic programming equations that provide a solution for an MPE. The study of the dynamics and evolution of an industry that incorporates the problem of selection has been carefully carried out by Jovanovic (1982). In this work, firms learn about their efficiency as they operate, and exit when they realize that their marginal costs are too high. This dynamic is also exhibited in our results.

Numerical solution to an MPE in which firms invest in R&D has been introduced by (Pakes and McGuire, 1992; Ericson and Pakes, 1995). Our simulations draw heavily from their algorithm. The first framework has been extended by many authors in the last 25 years, Doraszelski and Pakes (2007) describe the main contributions. Recently several papers have advanced the idea of prices as investments (see for example Besanko et al. (2010, 2014)); they develop a model with a learning-by-doing feature, and showcase firm's incentive to strategically decrease its prices in order to deny technological improvements to competitors. These papers focus on a limited number of competitors (typically two or three) that compete in each turn; such a restriction is due to the computational complexity of the MPE. However, modeling sellers' behavior in this way would misrepresent the real-world problem. In fact, sellers compete with a potentially large and uncertain number of players, whose identity changes from trip to trip. Weintraub et al. (2008, 2010) introduce a concept of an Oblivious Equilibrium which is shown to be a good approximation of an MPE, despite being computationally much less demanding.

Second, previously mentioned papers have explored R&D investments and a learning-by-doing mechanism to model industry evolution, our work focuses on the reputation of sellers as a state variable. The literature on the value and the impact of online reputation has emerged at the beginning of the 2000s, and many early papers were focused on eBay's reputation system. Livingston (2005) shows that sellers on eBay are strongly rewarded with the first few positive reviews; Cabral and Hortacsu (2010) also confirm these results, they also show that early negative reviews are of particular importance; Jolivet et al. (2016) study the impact of seller's reputation on equilibrium prices; notably, by a close attention to unobserved sellers heterogeneity, their approach is related to our reduced form results. Reputation has also been shown to matter in an offline set-up by for example Spagnolo (2012). This paper is particularly important for us because it models entry to the market with a reputation system. Finally, we assume that selles have a quality component that is initially unknown both to them and to the market. Over time, information about the sellers is revealed through reviews left by the passengers. Thus, we model sellers as experience goods (Nelson, 1970). Our paper is then related to the literature on dynamic pricing of experience goods. Bergemann and Välimäki (2006) analyze the problem of the monopolist selling a new experience good without any initial private information about its quality. They show that depending on the discount factor two price profiles can emerge: in "niche markets", the monopolist starts with low introductory prices and increases over time; in "mass markets" the opposite price profile arises. Bergemann and Välimäki (1996, 2000) analyze the interplay between learning and pricing in models with entrant and incumbent sellers. The entrant has no private information about its quality. Bergemann and Välimäki (1996) takes the case of a single long-lived buyer. If the entrant makes a sale in the first period, the revelation of information will induce higher prices in expectation in later periods. Thus, in all the "cautious" MPEs of the game, the entrant posts introductory prices that are lower than the marginal cost. The interplay between the increased competition for future market shares and the decreased price elasticity of demand of long-lived buyers is revisited in Miguel Villas-Boas (2006). In Bergemann and Välimäki (2000), buyers are short-lived thus very elastic to prices, the entrant makes even deeper discounts in the early periods in order to build a reputation. These papers focus on the level of experimentation that arises in the equilibrium: dynamic pricing leads to an efficient level of experimentation when the buyers are long-lived, but to too much experimentation when buyers are myopic. Hagiu and Wright (2018) extend these results to the context of a platform. Overall, when there is competition between sellers with no private information, theory predicts entrants post low and increasing prices in early periods.

3 Empirical model

We propose a model of dynamic oligopoly, where the sellers are long-lived, but they have to take a decision whether to trade or not in each period. If they decide to trade they have to forgo an opportunity cost, but they benefit from per-period profit and improve their continuation value by building reputation. Formally, the model is similair to Ericson and Pakes (1995), where all incumbents exit every period, and entry decisions are effective in the same period as when they are taken. Each period t we split into entry decisions, profits, and reputation updating stages. Our empirical strategy follows closely the literature on Experienced Based Equilibria (Pakes et al., 2007; Fershtman and Pakes, 2012).

3.1 Demand for ride-sharing services

We assume passengers looking for a ride from city A to city B to be comparing among all rides available during a given day, and choosing the one that maximizes their utility. Demand is then a standard discrete choice setup : Passenger i chooses among all the listings available on route r and day t of the trip the listing that maximizes their utility.

If passenger i chooses a listing by seller j, they would be getting the utility :

$$u_{i,j,r,t} = \alpha_1 p_{j,r,t} + \alpha_2 \mathbb{E}[q_{j,r,t}|I_{j,t}] + \alpha_3 n_{j,r,t} + \alpha_4 nl_{j,r,t} + \beta X_{j,r,t} + \epsilon_{i,j,r,t} \tag{1}$$

where $p_{j,r,t}$ is the price set by seller j when listing on route r and day t, $\mathbb{E}[q_{j,r,t}|I_{j,t}]$ is the perceived quality of seller j given the signal $I_{j,t}$ they have received up to time t, $n_{j,r,t}$ is a concave transformation of the number of reviews, and $nl_{j,r,t}$ is a concave transformation of the number of trips the driver has already made on the platform. $X_{j,r,t}$ is a collection of listing-specific variables, observable to both the passenger choosing a listing, and the econometrician. $\epsilon_{i,j,r,t}$ is i.i.d across passengers, sellers, route and day. Passengers can choose an outside option and get the utility $u_{0,r,t} = \epsilon_{i,0,r,t}$. We allow the outside option to vary by route and day. Conceptually, the outside option includes alternative modes of transport like train, or a choice of travel by ride-sharing on a different day.

One can be concerned with endogeneity of price in the variables. We have individual level data to use for estimation, and although per-se this does not alleviate concerns of price endogeneity; combined with the observability of all dimensions that are observed by the passenger at the time of the choice reduces the concerns for an omitted variable bias. Channels that could create endogeneity are through the textual description of the listing, whether the person is smiling in their photo or not, and the textual content of the last ratings.

3.2 Theoretical model of market entry

A large number M of long-lived potential sellers decides every period whether to trade or not. Each seller j is characterized by $x_{j,t} = (n_g, n_b, c)$ at time t. These characteristics include the reputation of the seller (number of good and bad signals they received up to the current period) and their marginal cost of providing the service c_j, all of them being non-negative, discrete numbers with upper bounds at $\overline{n}_g, \overline{n}_b, \overline{c}_j$; the state space Ω takes all potential combinations of $(n_{g,j,t}, n_{b,j,t}, c_j)$. Let Ξ_t denote a number of potential sellers, who decide to trade in a the period t; X_t is a vector that summarizes the number and characteristics of sellers who trade in period t; so for example the first entry is the number of sellers with no reviews (neither good or bad) and zero marignal cost. Let z_t be a vector of exogenous profit shifters, which evolves following a finite-state Markov process. We assume that per period profits of seller j are fully characterized by these variables: $\pi_{j,t}(x_{j,t}, X_{jt}, z_t; \theta)$, where θ is a paremeter vector, the true value of which is θ_0. Reputation of sellers is updated after profits are realized and it follows a probabilistic process; per period profits are increasing in the number of good reviews and decreasing in number of bad reviews. Furthemore, we assume no strategic behavior by sellers after the entry decisions are realized. Finally, the market structure X_t is determined by the profit function $\pi(\cdot)$ and the profit shifter z_t, thus for each value of z_t a potential seller can form an expectation of the number and characteristics of potential competitors.

Each of the potential sellers observes z_t and decides to be active if the expected profits in period t plus change in the value of continuation is higher then the opportunity cost. The opportunity cost is ϕ_{jt} and δ is the discount rate. Sellers problem is described by a Bellman equation:

$$V(x, X, z, \phi; \theta) = \max \left\{ \phi + \delta VC(x, z; \theta), \pi(x, X, z; \theta) + \delta VC(x', z; \theta) \right\}$$

,where x' denotes the expected update to the reputation of a seller. Sellers choose to be active when the latter term in the max problem is higher. $VC(\cdot)$ is the continutation value, which is the expectation of the next period realization of the value function.

$$VC(x, X, z, \phi; \theta) = \sum_{x',z',X} \int_{\phi'} V(x', X', z', \phi; \theta) p(d\phi'|\theta) p(X'|z, \chi = 1) p(z'|z) p(x'|x)$$

$p(\cdot)$ denotes the probability distribution: for example $p(x'|x)$ is distribution of sellers state in the next period x' condition on her being in the state x, and $p(X'|z, \chi = 1)$ is the expected market structure in the next period, condition on the market being in state z today and the seller being active next period ($\chi = 1$). As hinted by Pakes et al. (2007), in the equilibrium seller's expectation of the market structure in the future periods has to be consistent with the expectations of others sellers. We use following assumptions[2] to determine the equilibrium behavior.

Assumption 1: The distribution over the opportunity costs $F^{\phi}(\cdot|\theta)$ is nonnegative and generates costs i.i.d across time and markets. The distribution is known to all sellers, but the realizations are observed only by a given seller (asymmetric information).

Assumption 2: Perceptions of the market structure in period $t+1$, $p(X_{t+1}|z_t, \chi = 1)$ is influenced only by the profit shifter z_t. The evolution of the profit shifters follows Markov process, and $\pi(\cdot)$ is bounded.

With these assumptions we characterize an equilibrium, the key observation is that entry decisions are simultaneous based on the market conditions z that are obseved by everyone and opportunity costs.

Proposition 1. *An equilibrium is a collection of optimal entry decisions of each seller $\{e^*_{j,t}\}_{t\in 0,\infty}$ given their opportunity cost and profit shifters $z_{t\in 0,\infty}$. and a market structure $\{X_t\}_{t\in 0,\infty}$ such that:*

1. *All sellers have consistent perceptions of the likely market structure $p(X'|z, \chi = 1)$*
2. *The entry decision is individually rational for a seller in a state x_j with opportunity cost $\phi_{j,t}$, $e^*_{n,t}(\phi, z_t) = \mathbb{1}_{\pi(x_{j,t}, X_t, z_t;\theta) + \delta VC(x', z;\theta) - \phi_{j,t} - \delta VC(x,z;\theta)}$*
3. *The market structure $\{X\}_{t\in 0,\infty}$ arises with the optimal entry decisions of the sellers.*

[2]These are very similair to the assumptions in (Pakes et al., 2007; Fershtman and Pakes, 2012).

Perceptions of the future market structure must perceive that the probability of entering for sellers of a given state as identical up to their draws on opportunity costs. As long as the distribution of seller types in population M is known sellers can form expectations which will be consistent with equilibrium play. Let the state space Ω be indexed by k, with $K = Card\,[\Omega]$.[3]

$$p(X|z, \chi = 1) = \sum_{k=1}^{K} F^{\phi}\left\{VC(X, z; \theta)|\theta\right\}$$

Properties of equilibrium policy functions are discussed in details by the abovementioned literature on the Experienced Based Equilibria, a crucial observation that will allow us to write down the empirical counterpart of the value functions characterized above is that there is a set of recurrent market structure, profit shifters combinations (X, z), and in each of them the equilibrium play of the sellers gives rise to the same data generating process. If the dataset is large enough, so that all recurrent (X, z) combinations have been observed, and as a consequence the realized profits are known as well, we can calculate the continuation values.

3.3 Estimation strategy

The key methodologial challenge is to estimate the value of entering the platfrom from observed market outcomes. A seller takes into account impact of entering the market today on the likely update to her state, but also has to take into account probable future states of the market as well as her future draws of the opportunity cost. The simplification of this problem which we propose in this section stems from the claim that if the dataset is rich enough it encompasses all recurrent industry states: in our case this translates into profits of a seller x_j under different realization of (X, z), and the observed transition probabilities $p(x'|x)$.

The estimation strategy that we follow entails two steps:

1. Compute averages of the entry values (per period profits for sellers of any state) and transition values $p(x'|x)$, which is simply the probability of obtaining a good or a bad review in period $t+1$ conditioned on a number of good or bad reviews in period t. These together allow us to calculate expected profits conditioned on being active in future periods.

2. We take the estimated values from the Step 1 and treat them as actual expectation of sellers, to estimate the distribution of ϕ. For a given market we observe several realizations of (z, X),

[3]By assuming that composition of M is constant, so market is in some steady-state, we could potentially have an assumption about the share of sellers of different state in the total M; thus we can have the expected market structure

which together with our estimtes of $VC(x_j, X, z)$ allows us to characterize $f(\phi)$.

We can rewrite $VC(x, X, z, \phi; \theta)$ as:

$$\mathbf{E}_{\phi', z', x'} \left[\max \left\{ \phi' + \delta \mathbf{E} \left[VC(z'', x') | z', x' \right], \mathbf{E} \left[\pi(z', x') | z, x \right] + \delta \mathbf{E} \left[VC(z'', x'') | z', x' \right] \right\} \right]$$

,where notation a' denotes the realization of random variable a in the next period, and a'' two periods ahead. The maximization problem can be represented as the choice probability problem, to obtain:

$$VC(x, X, z, \phi; \theta) = \tag{2}$$

$$Pr \left\{ \phi < \mathbf{E} \left[\pi(z', x') | z, x \right] + \delta \Delta VC(z', x') \right\} \cdot \left[\mathbf{E} \left[\pi(z', x') | z, x \right] + \delta \mathbf{E} \left[VC(z'', x'') | z', x' \right] \right]$$
$$+ Pr \left\{ \phi > \mathbf{E} \left[\pi(z', x') | z, x \right] + \delta \Delta VC(z', x') \right\} \cdot \mathbf{E} \left[\phi' | \phi' > \mathbf{E} \left[\pi(z', x') | z, x \right] + \delta \Delta VC(z', x') + \delta \mathbf{E} \left[VC(z'', x') | z', x \right] \right]$$

,where $\Delta VC(z', x') = \mathbf{E} \left[[VC(z'', x'') | z'', x'] - [VC(z'', x') | z'', x'] \right]$, is the expected gain in value of being on the platform due to one review, where the expectation is with respect to probabilities of obtaining a good or bad reviews conditioned on being in state x'. For an illustration assume that the opportunity costs follow exponential distribution $F(\phi) = 1 - e^{-1/\sigma \phi}$, which implies that $\mathbf{E} [\phi' | \phi' > A] = A + \sigma$. Thus, we can simplify equation 2 to write that:

$$VC(x, z, \phi; \theta) = \tag{3}$$

$$\mathbf{E}_{z', x'} \left[\pi(z', x') + \underbrace{Pr \left[\phi' | \phi' > \mathbf{E} \left[\pi(z', x') | z, x \right] \right]}_{p(\phi')} \sigma + \delta \left(VC(x', z', \phi; \theta) + (1 - p(\phi')) \Delta VC(z, x) \right) \right]$$

Denote by M the matrix of transition probabilities as perceived by the sellers: such that element (i, j) denotes probability of transition from a state (z_i, x_i) to a state (z_j, x_j), so it captures demand process $z_i \to z_j$, as well as reputation building: $x_i \to x_j$, which entails probability of getting a high review or a low one. We can represent the equation 3 in the matrix form:

$$VC(\theta) = M [\pi + p(\phi')\sigma] + M [\delta (VC(\theta) + (1 - p(\phi')) \Delta VC(\theta))] \tag{4}$$

The number of good and bad reviews is bounded; we interpret this assumption that in practice it does not matter whether a seller collects more reviews, once she collected a stock of them. Thus, for each seller in state $l > L$, $\Delta VC(z, x_l) = 0$.

By plugging the defintion of $VC(\theta)$ into equation 4 and iterating we can get rid off the $VC(\theta)$ from the right hand side and obtain:

$$\begin{aligned} VC(\theta) &= \sum_{\tau=1}^{L} \delta^{\tau} M^{\tau} \left[\pi + \sigma p(\phi') + \delta(1 - p(\phi'))\Delta VC(\theta)\right] + \sum_{\tau=L}^{\infty} \delta^{\tau} M^{\tau} \left[\pi + \sigma p(\phi')\right] \\ &= \sum_{\tau=1}^{\infty} \delta^{\tau} M^{\tau} \left[\pi + \sigma p(\phi') + \delta(1 - p(\phi'))\Delta VC(\theta)\right] \end{aligned} \tag{5}$$

,where the last equality is obtained by noting that $\Delta VC(z, x_l)$ takes values of zero in a subset of states. Finally, the equation 5 can be further simplified to obtain:

$$VC(\theta) = \left[I - \delta M\right]^{-1} M \left[\pi + \sigma p(\phi') + \delta(1 - p(\phi'))\Delta VC(\theta)\right]$$

The first step in the estimation provides us with estimates of $\pi(z, x)$ for all (z, x), gain in profits $\Delta VC(\theta)$ is also estimated in the first step. Transition probabilies are directly observed in the data, the unknown parameter that has to be esimated is only σ.

4 Simulations of the theoretical model

In this section, we simulate solutions to the theoretical model. For the sake of simplicity, we introduce several assumptions that make the model easy to solve; however, still allow us to study how costs determine entry, exit, and pricing decisions. We are interested in two comparisons: first, if sellers are myopic in pricing, how their entry and exit decisions will depend on the distribution of costs, and how these will change if sellers start pricing dynamically (incorporate the impact of today's prices on tomorrow's profits). Second, what is the role played by the opportunity costs? How much heterogeneity in opportunity cost is needed to generate the exit patterns observed in market studies (for example JP Morgan Chase Institute).

We hold fixed a competitive setting : 3 sellers compete in a market. The demand follows a logit specification with elasticities of price and reviews. The game is dynamic in the sense that sellers build a reputation from period to period, and we allow for exit (exit decision are driven by comparing all expected future profits with the opportunity cost). Reputation follows a probabilistic transition rule; *ex-ante* sellers have the same quality, and their expected first grade is the same, the past realizations of the grades influence the future grades. Thus, we assume that sellers have an intrinsic quality that is

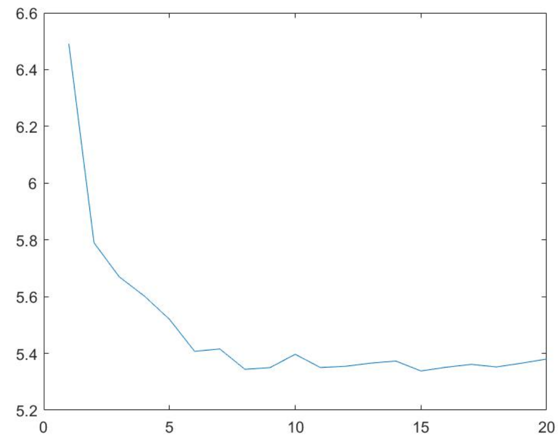

Figure 1: Average price in the market

revealed in time, it is, however initially unknown both to the seller and to the market.

4.1 Model with opportunity costs

First, we simulate solutions to the model with opportunity costs. Timing of the game is as follows:

1. Sellers observe their reputation, the reputation of their competitors, and **opportunity costs**, which are drawn at random each period. They decide whether to stay at the platform or exit. They exit when the discounted sum of future profits is lower than the opportunity costs. We do not allow for re-entry.

2. Sellers set prices and profits are realized.

3. Reputations of sellers who stayed are updated.

We performed 600 simulations of Nash solutions to the demand problem and exit decisions. Sellers set Bertrand prices and exit the market if the draw of opportunity cost is higher then the value of continuation. In this sense, the exit decisions incorporate the static-dynamic tradeoff, while pricing is myopic. Figures 1 and 2 show results.

Figure 1 shows the average market price. We see that the average price is initially high and gradually decreases. Close inspection of Figure 2 reveals what influences such behavior. Initially, sellers with a high marginal cost set higher prices; however, they leave the platform the moment they receive a high draw of the opportunity cost. Sellers with low marginal costs remain on the platform and start building a reputation, which allows them to start gradually increasing the prices.

In order to uncover the effect of reputation building on prices, we focus on sellers who stayed for at

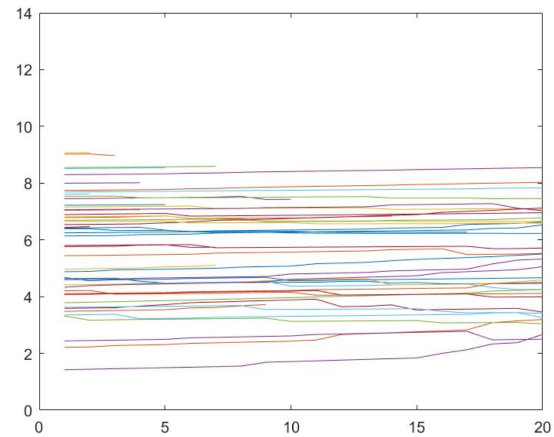

Figure 2: Example of evolution of prices: randomly selected markets

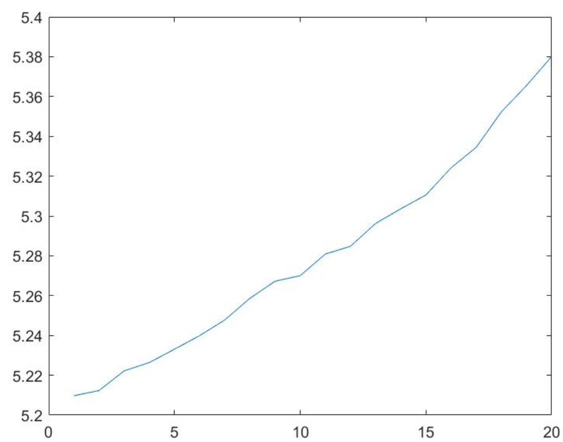

Figure 3: Average price of sellers who stayed for 20 periods.

least 20 periods. Figures 3 and 4 show the average prices by drivers who stayed pricing paths of some randomly selected drivers.

We draw two noteworthy conclusions from this simulation exercise: heterogeneity in marginal costs is crucial in generating price dispersion. Combination of opportunity cost and marginal costs shape exit patterns. Prices are initially high due to sellers with high marginal costs and decrease as they exit the market. Conditioned on staying on the market, a myopic entrant had set a lower price in the early periods than in the following periods - once she has build reputation, the optimal price is higher. Finally, observing changes in the distribution of prices and exit patterns, we can deduce the distribution of marginal costs and draws of opportunity costs.

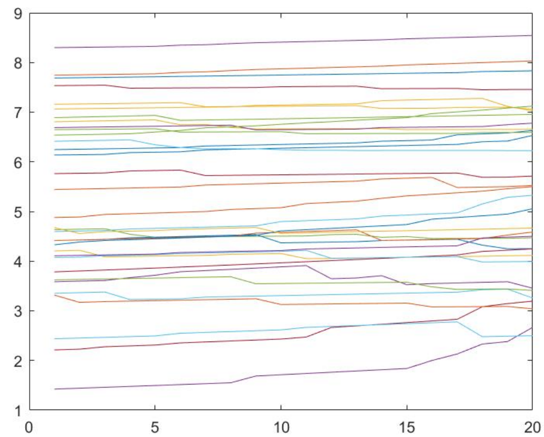

Figure 4: Prices of all sellers who stayed.

4.2 Model without heterogeneity in opportunity costs

In this subsection, we introduce a change to the baseline model: we remove heterogeneity in opportunity cost. This exercise should allow us to understand better the role played by opportunity cost. Sellers differ in marginal costs. The timing of a period is the following:

1. Sellers observe their reputation, the reputation of their competitors, **shocks to their marginal cost** which are drawn at random each period. They decide whether to stay at the platform or exit. They exit when the discounted sum of future profits is lower than the opportunity costs (the same for all sellers and constant over time). We do not allow for re-entry.

2. Sellers set prices and profits are realized.

3. Reputations of sellers who stayed are updated.

We simulate per period Nash prices, as set by myopic agents, and exit decision is a cut-off rule. Figures below show the results of 600 simulations.

Figures 5 and 6 show evolution of market prices (all sellers). We see that similar patterns emerge despite no heterogeneity in opprotunity cost. Prices are initially high and decrease gradually as inefficient sellers exit the market.

As in the previous simulations, we focus now on sellers who stayed for at least 20 periods. We observe the same patterns despite no difference in opportunity costs across sellers.

The second set of simulations reveals an interesting observation. Individual heterogeneity in opportunity costs is essential to understand individual entry and exit dynamics, however, average price dy-

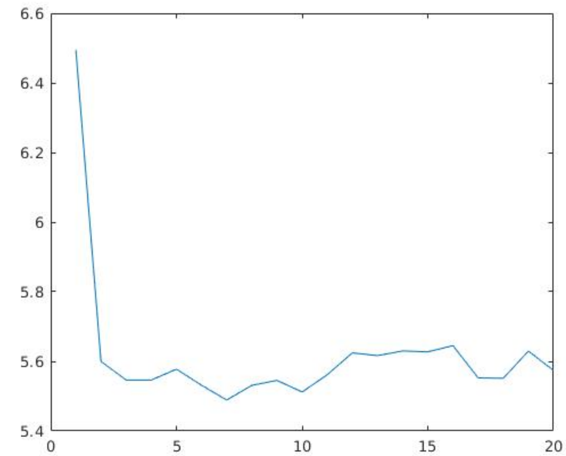

Figure 5: Average price in the market

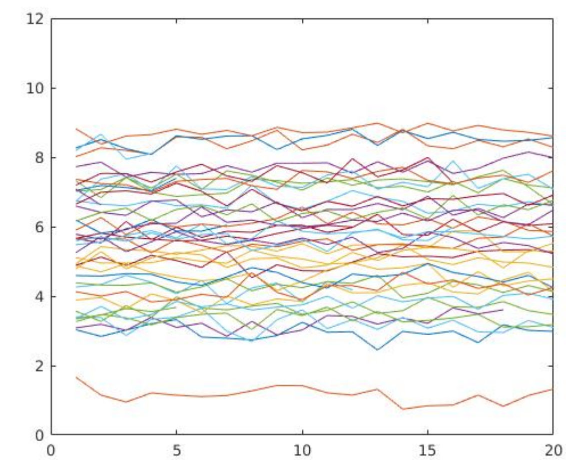

Figure 6: Example of evolution of prices.

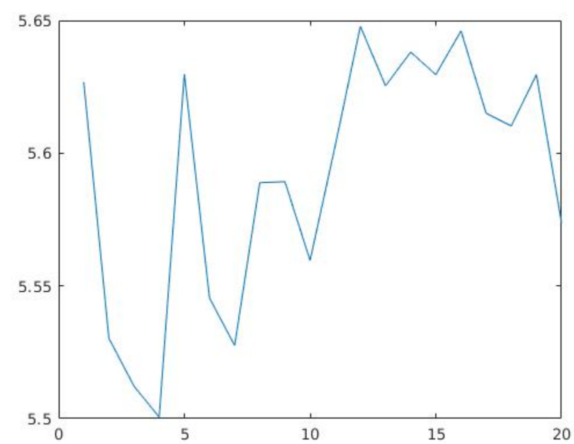

Figure 7: Average price of sellers who stayed for 20 periods.

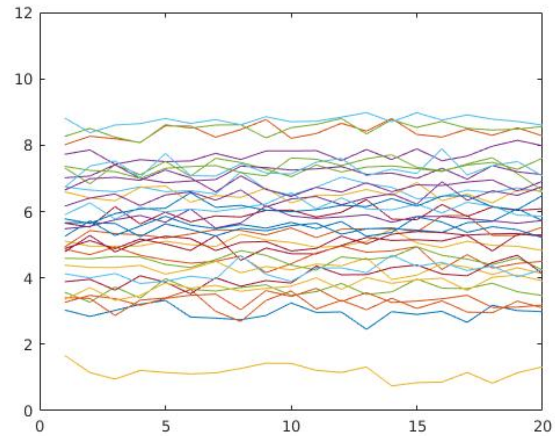

Figure 8: Prices of all sellers who stayed.

namics are shaped by the difference in marginal costs. Sellers with the high marginal cost set higher prices and are more likely to exit. Efficient sellers start with low introductory prices and increase them gradually as they receive reviews. These patterns emerge without dynamic pricing by sellers. Arguably, in a fully dynamic model, efficient sellers with including a further price reduction to speed up the reputation building process, which in equilibrium will result in the faster exit of inefficient sellers. We are currently working on incorporating this dynamic in our model.

5 Empirical application

The goal of this section is to test the predictions of our model with data from an online marketplace. First, we discuss the dataset that we use. Second, we provide some reduced-form statistics to show some patterns consistent with our predictions. Finally, we use the data to calibrate simulations of the theoretical model and to generate counterfactuals (*This section is underdevelopment*).

5.1 Data and empirical context

The dataset comes from Lambin and Palikot (2019); we are grateful to the authors for agreeing to share it. Here we briefly present the dataset, please refer to the original paper for the details of the collection, and processing. Table with summary statistics and the definition of main variables are in the Appendix.

Blablacar is a popular French ridesharing platform; it was established in 2006. Today it has around 60 million users worldwide; in France, it is used by 1.5 million travelers every month[4]. The platform

[4]https://blog.blablacar.fr/about-us/qui-sommes-nous

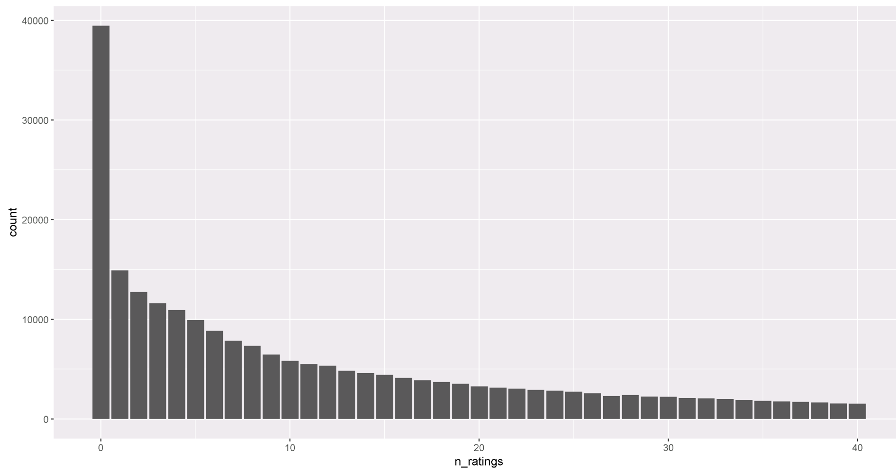

Figure 9: Share of drivers on different levels of experience.

caters to mostly non-professionals drivers looking to cover the costs of a trip from A to B. Blablacar experienced a period of fast growth in the first ten years of its functioning. However, regulatory changes of *Loi Macron*, in particular, liberalization of inter-city bus services have increased competition in the transportation market in France, and as a consequence lead to stagnation of Blablablacar in France.

There are some critical institutional differences on Blablacar's marketplace compared to other ridesharing platforms (e.g., UBER): drivers set their prices, and passengers choose a driver they want to travel with from a list of available alternatives (some drivers reserve a right to reject requests made by passengers). The decision to enter or not reflects the driver's expectation on demand conditions, and as consequence profits, she will be able to achieve. The key factor for our analysis is the composition of sellers on different levels of experience. Figure 9 shows that there are much more entrants to the platform then experienced users; this observation combined with the fact that Blablacar is not a growing marketplace, reveals an interesting observation: most of the entrants leave the platform soon after creating a profile. This observation is consistent with other studies (*see: JP Morgan Chase Institute*).

Furthermore, our theoretical model suggests that sellers on different levels of experience will price differently. If returns from reputation are sufficiently high prices of experienced drivers will be higher. On the contrary, is there is a significant amount of heterogeneity in opportunity costs, the composition of the sample will change as drivers exit: that might lead to more efficient drivers staying, and thus the average price would decrease. Figure 10 plots prices against the number of reviews suggests that more experienced drivers set lower prices than entrants. Figure 11 shows the distribution of prices for users

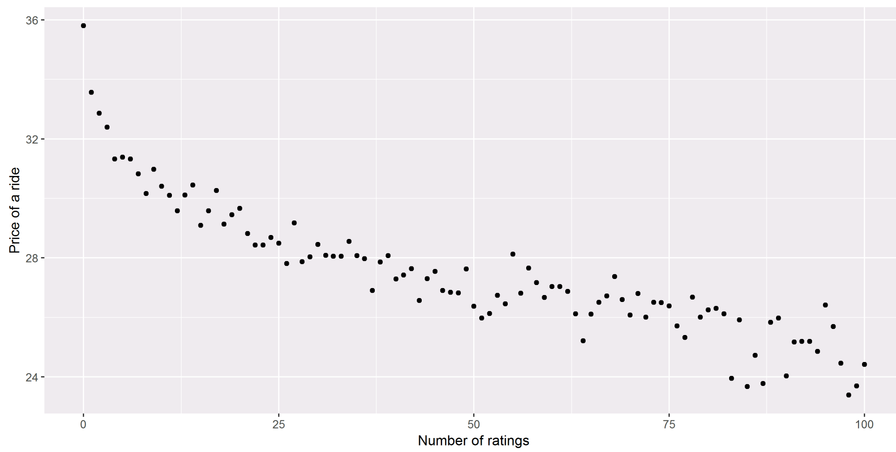

Figure 10: Mean prices on different stages of experience.

with no reputation, some reputation, and experienced ones. We see that prices set by drivers with no reviews are skewed to the left, and this skewness increases as drivers become more experienced. There is a mass of high prices that gradually disappears; later we will argue that the drivers who initially set high prices leave the platform.

At first, these patterns are quite surprising, seem to be contradictory with findings of the literature on dynamic pricing of experienced goods (Bergemann and Välimäki, 2000), and with prior empirical work on prices in online markets (Jolivet et al., 2016). However, these patterns are consistent with simulations of our model of section 3.

5.2 Empirical strategy

In this subsection, we use a reduced-form econometric model to, first, study the effect of early ratings on the pricing of drivers on the platform. Second, we relate pricing behavior, and ratings to the probability of exiting the platform *this point is under development*

Let $p_{j,r,t}$ denote the price posted by seller j, on a route r (pair of origin and destination cities) on day t. The econometric specification we use in this section follows the model

$$p_{j,r,t} = g(n_{j,t}) + q_{j,t} + \gamma X_{j,r,t} + \lambda_t + \lambda_r + \mu_j + \epsilon_{j,r,t}, \tag{6}$$

where price is a function of $n_{j,t}$ the number of past ratings a driver has accumulated at the time of the ride, $q_{j,t}$ the average of the ratings, $X_{j,r,t}$ time/route/driver characteristics that are observable to the econometrician, and time/route/driver fixed unobservable characteristics; μ_j denotes the driver-specific time-invariant characteristics that are relevant to the price-setting problem of the driver

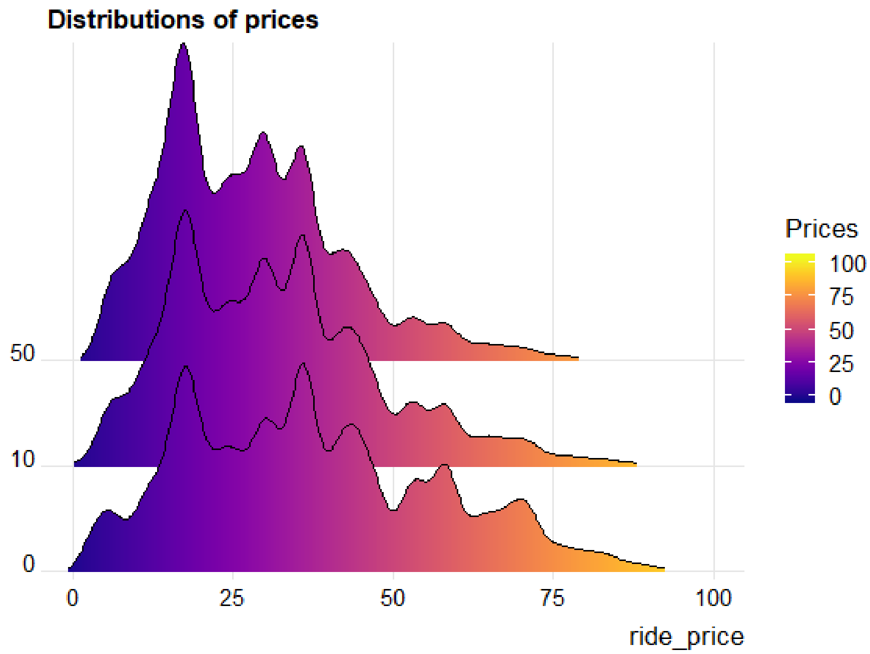

Figure 11: Distributions of prices on different stages of experience. Vertical axis number of reviews

but are unobserved to the econometrician; $\epsilon_{j,r,t}$ are idiosyncratic shocks that affect the pricing of the driver. We assume that the shocks $\epsilon_{j,r,t}$ are independent of the regressors at the time of the ride: $\mathbb{E}(\epsilon_{j,r,t}|q_{j,t}, n_{j,t}, X_{j,r,t}) = 0$

The function g characterizes the dependence of the seller's price on the number of ratings that she has collected before the time t. In practice, we estimate a non-parametric effect of binned ratings. The binning is thinner at early ratings and coarse for more experienced drivers: for example, the first bin consist of listings that have zero ratings, but all ratings above 100 constitute one bin.

Identification Our goal is to identify the effect of early ratings on sellers' prices. Are entrant sellers setting higher prices and decreasing them as they become more experienced, or are they starting with low introductory prices and increasing their prices gradually as they gain more reputation?[5]

The primary source of endogeneity that arises in our context is due to unobserved driver-fixed effects that determine both the price that the driver sets and the number of ratings she already has. For instance, drivers with high costs are less likely to post rides, and when they do, they set higher prices than their peers. These unobserved driver fixed effects appear to be the main source of bias in

[5]There might be other, non-costs, reasons for a seller to set a low introductory price, for example, Lambin and Palikot (2019) show in the same context that sellers from ethnic minorities have an incentive to decrease price upon entering the market in order to counter prejudice that they face.

our context.

Our identification strategy relies on including a rich set of observable characteristics and exploiting the panel-like structure of our data. First, in the data there is a measure of the "objective" travel cost of a route for a driver : it is computed as a function of the distance, the fuel consumption of the driver's car, the price of gasoline on the day of the trip in cities of departure and arrival and the highway tolls on the route traveled. The heterogeneity in route-day-specific travel cost, along with the heterogeneity in the driver-specific observable costs account for most of the variation in prices in our results below. We also include driver-specific controls (age, sex, automatic acceptance of the rider, ethnicity, etc.), and route/time specific controls (city, trip, the day of the week, strike day, etc.) Second, most of the drivers post several trips over their career, a sub-sample of these listings is observed for each driver. Thus, we can use within-driver variation in order to identify the effect of early ratings on the price set.

Another source of endogeneity is the potential correlation between the current number of ratings $n_{j,t}$ and past realizations of the idiosyncratic shock $\epsilon_{j,r,t}$. Although this is not an issue in the model in levels, the first-difference or within estimators which we use in order to exploit the within-driver variation might be biased if $E(\epsilon_{j,r,t}|n_{j,t'}) \neq 0$ for some t, t'. Informally, we think this bias may not be prevalent in our case as the first-difference and within estimates are similar: the bias created by such a correlation would result in a discrepancy between the first-difference and the within estimators. The difference in the estimates results from the fact that the two estimators use distinct methods to eliminate the fixed effect from the regression equation. The FD estimator bias would arise from $E(\epsilon_{j,r,t} - \epsilon_{j,r,t-1}|n_{j,t} - n_{j,t-1}) \neq 0$ whereas the within estimator bias would result from $E(\tilde{\epsilon}_{j,r,t}|\tilde{n}_{j,t}) \neq 0$, where $\tilde{n}_{j,t}$ is the within-driver demeaned number of ratings.

For a large share of drivers, we observe (1) more than two rides; and (2) rides posted at very different points of their careers (early on vs. later on). The effect of a past shock on today's number of ratings is decreasing with time, and we plan to use long-differences or Arellano-Bond style internal instruments to account for this source of endogeneity. In our analysis, we take the view that ratings are signals of the underlying quality of the driver. The noise in the past signals is independent of the past prices of the transaction, thus of the past idiosyncratic shocks : $\mathbb{E}(\epsilon_{j,r,t}|q_{j,t'}) = 0 \forall t, t'$: If ratings given by a customers depend on prices that they paid, we have a similar problem, which long-differences and internal Arellano-Bond instruments could address.

5.3 Results

In this section, we use a subset of rides spanning a year from July 2017 until August 2018, in total there are 302 502 observations. After controlling for the observables, we note that sellers with a higher number of reviews set lower prices than entrant sellers who do not have yet reputation. The average effect is negative and significant: an additional review is associated with a decrease of 3 cents of the euro (1% of the average price observed in the sample). In column (5) of Table 1, we report the estimates for the regression using the full sample and binned ratings. The effect is more accentuated at the early ratings: first four ratings reduce prices by 23 cents, moving from 4 to 16 ratings decrease prices by a further 37 cents; finally, ratings from 16 to 40 have a cumulative effect of 17 cents reduction in price. The effect is monotonously decreasing in the number of ratings.

As mentioned before, a unique aspect of the dataset that we use is that we have a useful measure of the objective or physical marginal cost of a ride for the driver. As we expect, the travel cost is the single most important variable in predicting the price; without any further controls, it explains around 80% of the variation in drivers' prices. From column (3) in Table 1, an increase of 1 euro in the travel cost is associated with an increase of 47 cents in the price of a ride on average. Controlling for trip fixed effects takes out the route-specific common cost between drivers, and the much lower estimates in column (4) and (5) now reflect the effect of between-driver heterogeneity. Drivers with highly efficient cars pass-through only a small fraction of this reduction in cost to their passengers (approx 2 cents for one standard deviation of fuel efficiency).

Many drivers appear in our dataset multiple times at various stages of their careers; hence, we can control for unobservable driver-specific fixed effects. We estimate two transformations of the price regression equation: a first-difference and a within-driver transformation. We include the binned number of ratings as before. In contrast to the OLS results, both estimators show a positive and significant impact of reputation on prices: therefore, once the unobserved driver characteristic is accounted for in the pricing decision, drivers with more ratings set on average higher prices. Collecting the first four ratings now *increases* the average price by 60 cents, going from 4 to 16 by a further 14 cents. The FD and within estimates are similar, and reflect a similar dynamic of pricing by the entrants: entrant sellers with no or little quality signals price lower than sellers with an established reputation.

As mentioned before, we expect a non-zero correlation between past shocks to pricing and current number (and quality) of ratings, but this correlation does not give rise to the same bias in the within and FD estimates. We do not formally correct for this bias. We expect doing so would decrease our

FD and within estimates of the effect of a rating, but we do not think it will change the dynamic of entrant's pricing.

Table 1: OLS on cross-section

	Dependent variable:				
	Price of a ride				
	(1)	(2)	(3)	(4)	(5)
Constant	39.822***	3.979***	1.490***	52.820***	54.200***
	(0.461)	(0.220)	(0.338)	(0.319)	(1.191)
1-2 reviews	−2.453***	−0.165***	0.051	−0.029	−0.121***
	(0.122)	(0.060)	(0.060)	(0.040)	(0.040)
3-4 reviews	−3.734***	−0.479***	−0.232***	−0.154***	−0.231***
	(0.130)	(0.063)	(0.063)	(0.041)	(0.041)
5-8 reviews	−4.779***	−0.709***	−0.485***	−0.309***	−0.378***
	(0.115)	(0.056)	(0.056)	(0.037)	(0.037)
9-12 reviews	−5.451***	−0.806***	−0.555***	−0.404***	−0.469***
	(0.129)	(0.062)	(0.062)	(0.041)	(0.041)
13-16 reviews	−6.029***	−1.084***	−0.812***	−0.577***	−0.607***
	(0.140)	(0.066)	(0.066)	(0.044)	(0.044)
17-20 reviews	−6.116***	−1.078***	−0.785***	−0.546***	−0.569***
	(0.152)	(0.071)	(0.071)	(0.047)	(0.047)
21-40 reviews	−7.599***	−1.580***	−1.262***	−0.807***	−0.773***
	(0.108)	(0.052)	(0.053)	(0.035)	(0.035)
41-100 reviews	−9.405***	−2.265***	−1.897***	−1.113***	−0.961***
	(0.107)	(0.052)	(0.052)	(0.035)	(0.036)
101+ reviews	−12.998***	−2.765***	−2.356***	−1.326***	−1.043***
	(0.117)	(0.057)	(0.057)	(0.038)	(0.039)
Average rating	−0.809***	0.096**	0.060	−0.199***	−0.339***
	(0.099)	(0.046)	(0.046)	(0.030)	(0.030)
Travel cost		0.478***	0.474***	0.022***	0.024***
		(0.0005)	(0.0005)	(0.001)	(0.001)
Travel Cost		X	X	X	X
Time effects			X	X	X
Trip effects				X	X
Driver Controls					X
Observations	298,185	238,509	238,509	238,509	235,082
R^2	0.054	0.832	0.835	0.929	0.931
Adjusted R^2	0.054	0.832	0.835	0.928	0.930
Note:					*p<0.1; **p<0.05; ***p<0.01

Results presented in Table 2 indicate that the unobserved heterogeneity plays an important role. This is also a price dynamic consistent with predictions of our theoretical model. We argue that this unobserved factor is constituted of a driver-specific opportunity cost of offering a ride on the platform. Such a cost includes committing to doing the ride and forgoing other opportunities.

The presence of heterogeneity in opportunity costs is consistent with the sign of the bias in the OLS estimates in Table 1. A driver with a higher average opportunity cost is less likely to post rides

Table 2: Repeated observation sample

	Dependent variable:		
		Price of a ride	
	Pooling	First-Difference	Within
	(1)	(2)	(3)
Constant	55.994***	−0.016	
	(1.514)	(0.012)	
1-2 reviews	−0.075	0.358***	0.404***
	(0.051)	(0.086)	(0.078)
3-4 reviews	−0.164***	0.531***	0.587***
	(0.052)	(0.090)	(0.079)
5-8 reviews	−0.309***	0.691***	0.738***
	(0.048)	(0.093)	(0.078)
9-12 reviews	−0.375***	0.692***	0.816***
	(0.052)	(0.102)	(0.084)
13-16 reviews	−0.560***	0.710***	0.740***
	(0.055)	(0.111)	(0.090)
17-20 reviews	−0.461***	0.778***	0.763***
	(0.058)	(0.119)	(0.095)
21-40 reviews	−0.721***	0.623***	0.667***
	(0.046)	(0.121)	(0.094)
41-100 reviews	−0.914***	0.338**	0.443***
	(0.045)	(0.136)	(0.107)
101+ reviews	−1.012***	0.133	0.057
	(0.047)	(0.167)	(0.129)
Average rating	−0.399***	−0.165*	−0.143*
	(0.034)	(0.099)	(0.086)
Travel cost	0.025***	0.034***	0.037***
	(0.002)	(0.004)	(0.003)
Travel cost	X	X	X
Time effects	X	X	X
Driver Controls	X	X	X
Trip effects	X	X	X
Observations	165,205	113,131	165,205
R^2	0.935	0.894	0.893
Adjusted R^2	0.935	0.893	0.844
Note:			*p<0.1; **p<0.05; ***p<0.01

on the platform, thus would have a lower number of ratings than a driver with a lower opportunity cost; if the higher opportunity cost is correlated with higher marginal costs, then it would also be associated with rides at a higher price: when the driver actually posts a ride, it is likely that the ride will be at a higher price than its peers. Although we think of the opportunity costs to be sunk at the time of posting, the pricing and posting decision are made simultaneously in our view. The unobserved heterogeneity in opportunity costs would then create a negative correlation between the number of ratings and the price, negatively biasing the OLS estimates. This is what observe in Table 2, the pooling estimator (Column 1) indicates a negative effect of the number of early ratings on price, but accounting for the unobserved heterogeneity, the effect of early ratings on price becomes positive (Column 2 and 3).

Second, we provide another robustness check by restricted the dataset for drivers that stayed on the platform long enough to receive 40 reviews, and that were observed in the dataset when they had less than five reviews. This can be seen as selecting the subsample of drivers with low opportunity costs: in this subsample, the bias of OLS disappears, which is consistent with our interpretation of heterogeneity in opportunity costs causing the bias in prices. Details of this robustness check are in the Appendix.

Third, another unique feature of the dataset is the fact that we observe exits by drivers. Each driver on Blablacar has her unique ID; thus, we can revisit profiles of drivers in order to check how active they have been after we have observed them. This allows us to establish whether drivers with high objective marginal costs are more likely to exit, and what is the impact of the reputation system on exit.[6] Exploiting the exit data, we note that drivers who have less than four new ratings in December 2018 compared to the last ride we observe in our dataset (up to August 2018) set significantly higher prices than those who have more than 4 new ratings (that is posted at least one more time since their last observed ride).

The three arguments presented above support a claim that the differences in opportunity costs are key factors in understanding the changes in the composition of the population of drivers, as well as their individual pricing decisions. They are all consistent with the theoretical model presented in section 3.

[6]We are currently developing this analysis; here we provide some intuitions.

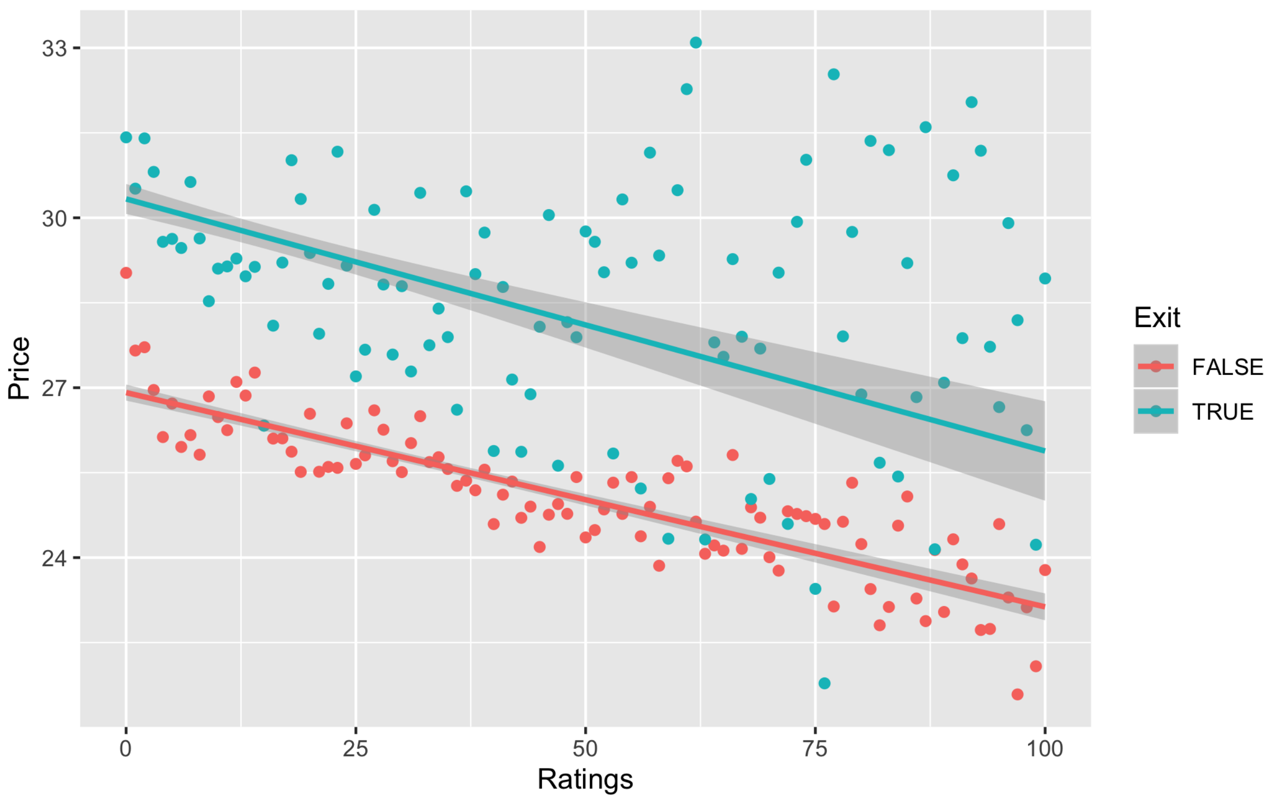

Figure 12: Prices of drivers who had posted at least 1 new listing by January 2019 since their last observed listing (in pink) are consistently lower than the prices of drivers who did not post a new listing since their last ride (in blue).

6 Results from the structural model

In the demand estimation, the market size $m_{r,t}$ is taken equal to the maximum number of views a listing generates in a route-day. The estimation is conducted on 10 routes including all available days. The estimates in the table below show that passengers are price-sensitive with an average price-sensitivity of -1.38.

Table 3: Conditional logit demand estimates

	Dependent variable:
	choice
Price	-0.05^{***} (0.004)
Ratings (log)	0.32^{**} (0.14)
Listings (log)	-0.14^{***} (0.02)
Average Rating	0.02 (0.08)
Ratings (log) x Average Rating	-0.01 (0.03)
Observations	2,594,650
Wald Test	$228{,}872.20^{***}$ (df = 49)
Note:	*p<0.1; **p<0.05; ***p<0.01

Using the estimated demand model, the observed market sizes, we recover sellers' marginal costs assuming full information per-period Nash equilibrium in prices, and in a first pass, static pricing. Prices set by sellers on the platform are per seat, so if the per-market Bertrand-Nash assumption holds, we can recover the marginal cost (per passenger) of a driver from the FOC of the seller problem :

$$\hat{c}_{j,rt} = p_{j,rt} + \frac{D_{j,rt}}{\partial D_{j,rt}/\partial p_{j,rt}} := c_r + \xi_j + w_{jrt}$$

We then use the estimated marginal costs (per passenger) to derive a simple test of the validity of the seller conduct assumption. The intuition of the test is simple : Under Bertrand-Nash with full-information (no learning) and myopic agents, the shock to the marginal cost of seller is uncorrelated to either the number of listings a seller has posted, or the number of reviews they have received up to period t : $E(w_{j,r,t} n_{j,r,t}) = 0$ and $E(w_{j,r,t} nl_{j,r,t}) = 0$. Findings are in Table 5. The estimated cost shocks exhibit some correlation with the number of trips listed by the driver, an indication about the presence

Within driver	0.61
Across drivers within route	3.62
Across drivers Across route	13.89

Table 4: Variance decomposition

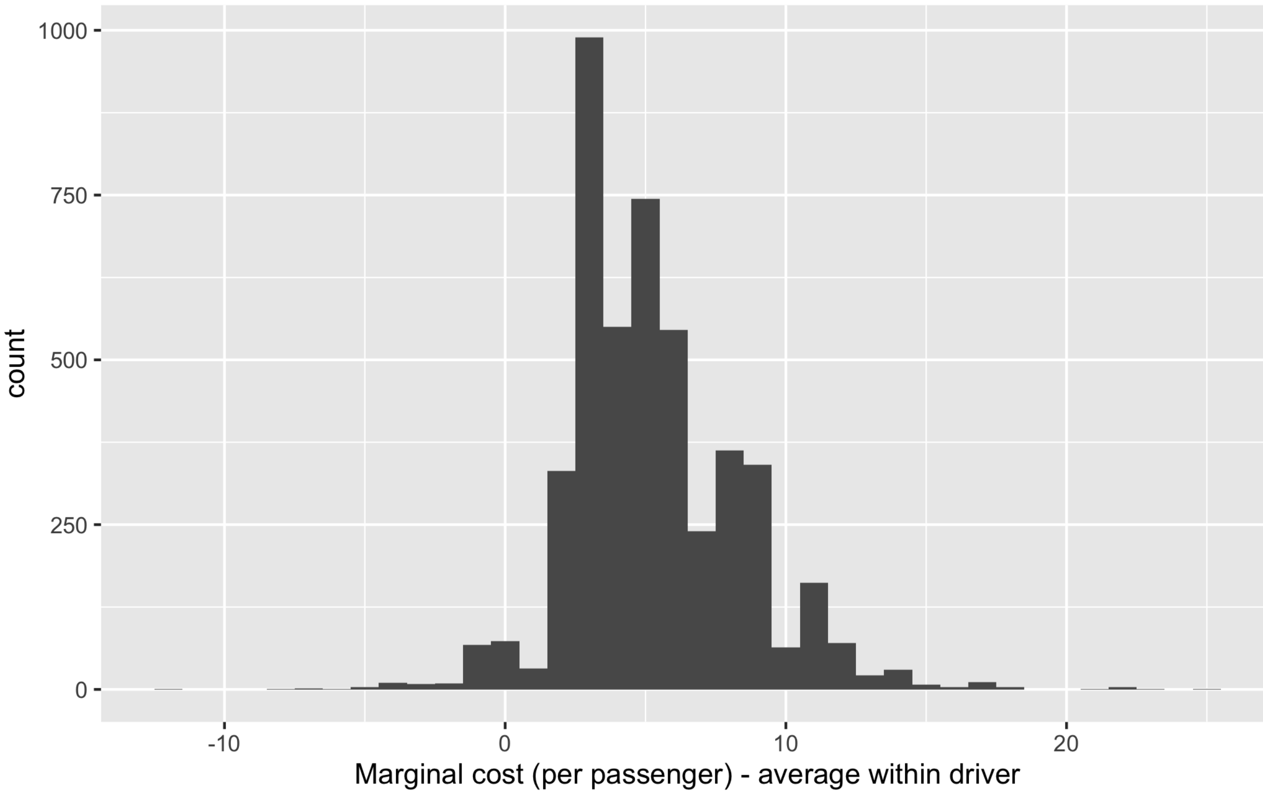

Figure 13: Costs of drivers on Paris-Angers

of learning.

Table 5: Marginal cost regressions

	Dependent variable:	
	Marginal cost (per passenger)	
	Pooling	Within
	(1)	(2)
Ratings (log)	-0.504^{***} (0.059)	0.190 (0.124)
Listings (log)	-0.170^{***} (0.065)	0.517^{***} (0.162)
Average Rating	0.080^{***} (0.011)	-0.019 (0.020)
Route effects	X	X
Weekend, SNCF strike	X	X
Driver unobservable FE		X
Observations	9,555	9,555
Adjusted R^2	0.951	0.758
Note:	*p<0.1; **p<0.05; ***p<0.01	

In order to substantiate our previous claims on the selection taking place on the platform, we offer some descriptive results on the correlation between marginal costs and the probability of exit. The estimated marginal costs as shown in the table below are a strong predictor of the probability of exit of a driver. A higher marginal cost driver is less likely to make a new listing on the platform even after controlling for the probability of sale at that price.

Table 6: Probability of exit after n trips and marginal cost

	Dependent variable:					
	Exit					
	(1)	(2)	(3)	(4)	(5)	(6)
Marginal cost (per passenger) c_j		0.041^{***} (0.007)		0.041^{***} (0.007)	0.022^{***} (0.007)	0.022^{***} (0.007)
No sales	0.158^{***} (0.057)	0.119^{**} (0.058)			0.008 (0.062)	
Revenue			-0.004^{***} (0.001)	-0.004^{***} (0.001)		-0.001 (0.001)
Average rating	0.194^{*} (0.105)	0.202^{*} (0.105)	0.199^{*} (0.105)	0.202^{*} (0.105)	0.004 (0.101)	0.004 (0.101)
Ratings (log)					-0.667^{***} (0.022)	-0.666^{***} (0.022)
Route effects	X	X	X	X	X	
Observations	9,546	9,442	9,468	9,442	9,442	9,442
Log Likelihood	$-4{,}955.883$	$-4{,}878.708$	$-4{,}904.485$	$-4{,}876.714$	$-4{,}349.097$	$-4{,}348.812$
Akaike Inf. Crit.	9,935.767	9,783.416	9,832.971	9,779.427	8,726.195	8,725.624
Note:						*p<0.1; **p<0.05; ***p<0.01

In this rest of this section, we wish to estimate and calibrate the supply side theoretical model presented in section 3 in order to study potential policy response by the market designer. We are particularly interested in studying the potential of the platform to respond to high exit rate by promoting different types of drivers.

7 Conclusion

In this paper, we propose a model to study entry, exit, and pricing decision of sellers in a marketplace with a reputation system. We show that the heterogeneity in marginal and opportunity costs is the key to understanding the evolution of prices, as well as exit decisions. We, first, provide simulations of our model and, second, show that reduced form analysis of pricing on a popular ride-sharing platform has to account for the unobserved heterogeneity, to uncover expected price patterns.

Sellers with high opportunity cost set on average higher prices, and are more likely to leave the platform. Sellers with lower opportunity costs initially set low prices, and gradually increase them when they have built a reputation.

We are currently working on extending this draft in two dimensions. First, as mentioned before, a unique feature of our empirical set-up is that we can revisit profiles of drivers and measure whether they are still active or decided to stop using the platform. We want to extend our reduced form results by a study of exit decisions. Initial results reinforce our previous findings: sellers with high marginal costs (high travel costs) are more likely to leave the platform; also, a negative review has an effect of increasing the probability of exit. Second, our supply-side model rests on a number of assumptions that do not match the reality of a ride-sharing platform well. We are working on extending the model and our simulations to allow for a large and changing number of competitors and their strategic entry decisions.

Finally, during the period at study, Blablacar introduced a policy of promoting entrants by granting higher prominence to their listings. The effects of such a policy depend on how sellers make their pricing decisions and the distribution of unobserved costs. Our structural model allows us to estimate the effects of this policy change.

References

Bergemann, D. and Välimäki, J. (1996). Learning and Strategic Pricing. *Econometrica*, 64(5):1125–1149.

Bergemann, D. and Välimäki, J. (2000). Experimentation in Markets. *Review of Economic Studies*, 67(2):213–234.

Bergemann, D. and Välimäki, J. (2006). Dynamic Pricing of New Experience Goods. *Journal of Political Economy*, 114(4):713–743.

Besanko, D., Doraszelski, U., and Kryukov, Y. (2014). The economics of predation: What drives pricing when there is learning-by-doing? *American Economic Review*, 104(3):868–97.

Besanko, D., Doraszelski, U., Kryukov, Y., and Satterthwaite, M. (2010). Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica*, 78(2):453–508.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Doraszelski, U. and Pakes, A. (2007). A framework for applied dynamic analysis in io. *Handbook of industrial organization*, 3:1887–1966.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of economic studies*, 62(1):53–82.

Fershtman, C. and Pakes, A. (2012). Dynamic games with asymmetric information: A framework for empirical work. *The Quarterly Journal of Economics*, 127(4):1611–1661.

Hagiu, A. and Wright, J. (2018). Platforms and the exploration of new products. *Management Science (forthcoming)*.

Jolivet, G., Jullien, B., and Postel-Vinay, F. (2016). Reputation and prices on the e-market: Evidence from a major french platform. *International Journal of Industrial Organization*, 45:59–75.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society*, pages 649–670.

Lambin, X. and Palikot, E. (2019). The impact of online reputation on ethnic discrimination.

Livingston, J. A. (2005). How valuable is a good reputation? a sample selection model of internet auctions. *Review of Economics and Statistics*, 87(3):453–465.

Maskin, E. and Tirole, J. (1988a). A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs. *Econometrica: Journal of the Econometric Society*, pages 549–569.

Maskin, E. and Tirole, J. (1988b). A theory of dynamic oligopoly, ii: Price competition, kinked demand curves, and edgeworth cycles. *Econometrica: Journal of the Econometric Society*, pages 571–599.

Miguel Villas-Boas, J. (2006). Dynamic Competition with experience goods. *Journal of Economics & Management Strategy*, 15(1):37–66.

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 78(2):311–329.

Pakes, A. and McGuire, P. (1992). Computing markov perfect nash equilibria: Numerical implications of a dynamic differentiated product model.

Pakes, A., Ostrovsky, M., and Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *the RAND Journal of Economics*, 38(2):373–399.

Spagnolo, G. (2012). Reputation, competition, and entry in procurement. *International Journal of Industrial Organization*, 30(3):291–296.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2008). Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2010). Computational methods for oblivious equilibrium. *Operations research*, 58(4-part-2):1247–1265.

Appendix A

name of a variable	description
price	price set by the driver in EUR; has to be lower than maximum price: 0.082 per km
suggested price	price suggested by Blablacar: 0.065 per km
age	age of the driver in years
reviews	number of reviews received by the driver
male	gender defined based on photo recognition and name
minority	takes the value of one when the driver is of Arabic or African origin, and zero otherwise; defined based on photo recognition and name (see. Lambin& Palikot (2019)) for details)
picture	takes the value of one when driver added a picture, and zero otherwise
talkative	categorical variable (bla, blabla, blablabla) indicating how talkative the driver is
bio	number of words in driver's description
ride description	number of words in ride's description
reputation	mean of grades received by the driver
published rides	number of rides ever published by the driver
number of clicks	number of clicks a given listing has received; clicking is necessary for booking a ride but not sufficient; measured at the moment of data collection
sold seats	number of seats already sold; measured at the moment of data collection
revenue	sold seats multiplied by price
posts per month	mean number of listings posted by the driver since she joined the platform
seniority	number of months since the driver joined the platform
competition	number of listings available on the same day on the same route
median revenue	mean of median revenues in cities of departure and arrival; source: INSEE
public transport	travelling time by public transport on the route at listings' departure time; source: Google API
train strike	SNCF official strike implicating a given route
value of car	price of a comparable car model in thousands of EUR; when a model of a car is not available mean price of a brand; source: ebay (scrapped data)
fuel consumption	mean fuel consumption of a model of a car; when model of a car is not available mean consumption of a brand; source: ADEME
length (km)	distance in km between cities of departure and arrival; souce: Google API
lengh (hours)	estimated driving time by a car on a given route and time; source: Google API
hours until departure	number of hours between data collection and a ride departure
posted since	number of hours between the posting of the listing and data collection
automatic acceptance	takes the value of one if booking requests are automatically accepted and zero if the driver chose to accept/reject requests manually
to fuel price	average price of a litre of diesel in a city of arrival in cents
from fuel price	average price of a litre of diesel in a city of departure in cents
toll viamich	total toll costs on a given route in EUR; source: https://www.viamichelin.com/
travel costs	mean of fuel costs multiplied by fuel consumption plus toll fees
weekday	takes a value of 1 on weekdays and zero on weekends
pets	takes a value of 1 if the driver accepts pets and zero otherwise
music	takes a value of 1 if the driver listens to music in the car and zero otherwise
smoke	takes a value of 1 if the driver accepts smoking in the car and zero otherwise
detour	categorical variable: 1 if no detour, 2 if some detour (up to 15 min), and 3 if more than 15 minutes detour
luggage	categorical variable: 1 if no luggage, 2 if small bags, 3 if big bags are allowes

Table 7: Definition of main variables

Table 8: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Price	211,164	26.979	12.933	5	80
Number of ratings	211,164	49.212	100.644	0	1,579
Number of listings	211,164	70.005	146.262	1	2,895
Average rating	211,164	4.595	0.275	1.000	5.000
Seats offered	211,164	2.649	0.767	1	4
Seats taken	211,164	0.327	0.634	0	4
Weekend	211,164	0.325	0.468	0	1
SNCF strike	211,164	0.042	0.201	0	1
Automatic acceptance	211,164	0.462	0.499	0	1

Table 9: Summary statistics- drivers

Statistic	N	Mean	St. Dev.	Min	Max
Last observed number of ratings	51,057	41.417	64.006	0	1,573
Last observed number of listings	51,057	54.460	89.588	1	2,104
Last observed average rating	51,057	4.591	0.253	1.000	5.000
Driver observations	51,057	2.104	3.027	1	117
Photo	51,057	0.883	0.321	0	1
Age	51,057	37.610	13.553	18	102
Male	51,057	0.688	0.463	0	1
Exit (0 new trips)	51,057	0.267	0.442	0	1
New trips	51,057	9.526	18.272	0	633

Table 10: Summary statistics - route/day

Statistic	N	Mean	St. Dev.	Min	Max
Number of listings	27,866	7.578	11.807	1	203
Mean price	27,866	30.036	14.325	5.000	80.000
SD price	20,557	2.991	2.393	0.000	42.426
Market size	27,866	48.772	58.381	0	1,398
Seats offered	27,866	20.073	31.272	1	565
Seats taken	27,866	2.476	5.255	0	133
Listings w/ 0 booked seats (share)	27,866	0.814	0.251	0.000	1.000
Listings w/ less than 5 reviews (share)	27,866	0.359	0.348	0.000	1.000
Listings w/ 1st trip (share)	27,866	0.130	0.266	0	1

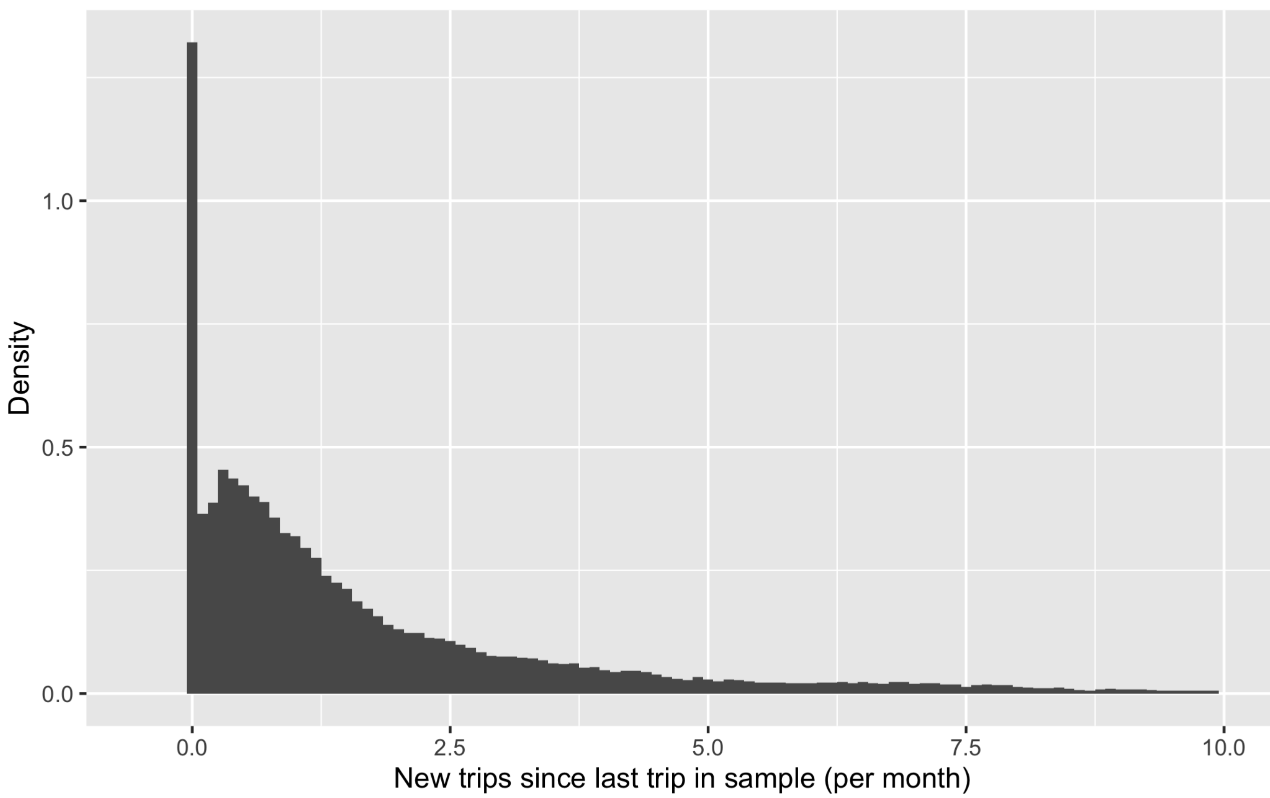

Figure 14: Distribution of number of new trips since last scraped listing (normalized per month).

Appendix B

Robustness check for the sign of the bias: we restrict our dataset to drivers whom we had seen when they entered the platform (five reviews or less) and stayed until they collected at least 40 reviews. This sub-sample is smaller than the unrestricted one; thus we estimate the effect of a coarser measure of being an entrant: having 5 or fewer ratings. The results are reported in Table 11. The within estimates of the pricing behavior of entrants (in column 4) shows a is a negative and significant sign, although there is a loss in precision. However, what is interesting is that in the subsample of drivers whom we observe on different stages of their careers, the pooling estimator of entrant status is also negative and significant. In this subsample, sellers price lower in early posts than later on, even without controlling for unobserved heterogeneity. Drivers in this subsample stayed on the platform for a long period; thus the selection of the subsample is not random: it is conditional on the driver not exiting before having at least 40 reviews. This observation indicates that conditioning on not exiting, the correlation between the unobserved fixed effect and the number of ratings becomes close to zero. We formally test this by a Hausman test of the within estimator versus the random estimator. We cannot reject the null hypothesis that both estimators are consistent in this subsample, whereas we reject it at 1% risk level in the larger sample. The unobserved driver-specific fixed effects that are driving our bias in Table 1 are then highly correlated with exit behavior. This gives credence to our interpretation of these effects as the average opportunity cost of the driver of using the platform: drivers with high opportunity costs leave the platform early on in their career but are still selected in the repeated observation sample of Table 2, thus the difference between the random and within estimators. The subsample we select for Table 11 amounts to conditioning on non-exit, this means selecting the drivers with low average opportunity costs. In this subsample, the random effects estimator is very close then to the fixed effects estimator: these are drivers with low opportunity costs who start at a lower price and increase gradually.

Table 11: Sub-sample with drivers we observe early and late stage of career

	Dependent variable:			
	Price of a ride			
	OLS	*panel*		
	Pooling	Within	Pooling	Within
	(1)	(2)	(3)	(4)
Constant	56.841***		−6.328***	
	(1.188)		(2.143)	
Entrant	0.588***	−0.427***	−0.711***	−0.209*
	(0.019)	(0.043)	(0.232)	(0.118)
Average rating	−0.275***	0.009	1.973***	−0.226
	(0.030)	(0.085)	(0.312)	(0.341)
Travel cost	0.030***	0.057***	0.417***	0.052***
	(0.001)	(0.003)	(0.002)	(0.008)
Sample	Full	Repeat observations	Early-late	Early-late
Travel cost	X	X	X	X
Time effects	X	X	X	X
Driver Controls	X	X	X	X
Trip effects	X	X	X	X
Observations	238,397	168,520	9,342	8,712
R^2	0.930	0.891	0.802	0.917
Adjusted R^2	0.930	0.842	0.802	0.905
Residual Std. Error	4.122 (df = 238111)			
Note:				*p<0.1; **p<0.05; ***p<0.01