Does reputation hinder entry?

Study of statistical discrimination on a platform *

Work in progress

Xavier Lambin[†]& Emil Palikot[‡]

September 30, 2017

Abstract

This paper studies entry dynamics of users facing statistical discrimination into a two-sided market with moral hazard. Firstly, we develop a theory model, where we study effort dynamics of users coming from different populations. We show that level of required effort and the expected surplus of agents are functions of their population. We also discuss when and why a platform will decide to create a reputation system. Secondly, we bring our theoretical results to data collected on a popular ride-sharing website. We show that minority male users face statistical discrimination during first interactions. However, the reputation system resolves this issue in the longer run.

1 Introduction

In many two-sided markets moral hazard plays an important role. Agents are often dependent on actions, taken by players on the other side of the market, that are not part of a contract. In the words of Joe Gebbia co-founder of AirBnB[1] a crucial element of success of this platform is *designing trust*. What he means by this is that AirBnB matches people who otherwise do not know each other, to engage in activity that has space for moral hazard. In particular, guests interact with hosts typically once, and both parties can exert effort that will make the other side of the platform better off. Such a platform could propose a screening contract in which participants of a high enough type will participate and

*We are grateful to Yassine Lefouili, Steven Tadelis and Bruno Jullien for their valuable comments on various stages of the paper.

[†]Ph.D candidate at Toulouse School of Economics and ENGIE , xavier.lambin@yahoo.com

[‡]Ph.D candidate at Toulouse School of Economics, emil.palikot@gmail.com

[1]https://www.youtube.com/watch?v=16cM-RFid9U

exert a lot of effort; other users would decide to opt out (see Roger and Vasconcelos (2014)). However, screening contracts that make low type agents not participate, can be suboptimal e.g. induce a low participation. Furthermore, statistical discrimination of particular minorities is a well-documented phenomena on many platforms labelling themselves as sharing economy (see for example Edelman and Luca (2014)). This sort of discrimination can make a simple screening contract difficult to implement or not optimal from the profit maximizing point of view. Development of a reputation system is often seen as a way out of this stalemate. On one hand, reputation system can serve as a commitment device allowing users to signal their type and commit to taking a high action, in exchange for higher probability of being matched. On the other hand, possibility of signalling one's type allows users facing initial discrimination to escape such a trap.

A goal of this project is to study implications of a reputation system on entry to a two-sided market suffering from a moral hazard problem. In particular, we are interested in dynamics of effort of users coming from different populations, including minorities facing statistical discrimination. In order to address this questions, we first develop a theoretical model, and secondly, we bring the hypotheses stemming from theory to data. We collect data through a web-crawling program on a popular ride-sharing platform. Our preliminary results point towards a claim that reputation system is an effective device in escaping statistical discrimination. Users coming from discriminated populations have to exert higher effort at the beginning, but after some time they are able to form a high reputation and escape such a discrimination. We show that minority male users are facing initial discrimination, they set prices on average lower, controlling for the quality of the car and outside options available on the route they travel. However, after several interaction i.e. once users have formed reputations, belonging to this minority does not longer play a significant role.

Relation to literature: Our projects falls into intersection of several strands of economic literature. Firstly, into the literature on two-sided markets, where the most relevant references being Rochet and Tirole (2003) and Caillaud and Jullien (2003). Our model can be seen, on one hand, as a special case of the former paper, where we focus on only two types of agents on each side of the market, on the other hand, as an extension as we also enrich the model with long-lived agents on one side of the market, and the moral hazard question. Problem of a long-lived oligopolist is studied by Maskin and Tirole (1988), and later on extended by Fudenberg and Tirole (2000), where similar feature of foreclosing the market by strategic shift by an incumbent is identified. Secondly, we consider a model of moral hazard, which is related to seminal works in the field Baron and Myerson (1982) and Laffont and Tirole (1986), but also more recent works of Roger and Vasconcelos (2014) and Garrett and Pavan (2012). Thirdly, recent economic (and computer science) literature has studied effectiveness and design of reputation systems, some of notable projects being: Nosko and Tadelis (2015), Cabral et al. (2010), Bar-Isaac and Tadelis (2008), Liu and Skrzypacz (2014), Livingston (2005), Jolivet et al. (2016), Bolton et al. (2004), Mayzlin et al. (2014), Jullien and Park (2014) and Zervas et al. (2015). However, these papers focus on understanding how consumers may react to the information provided. They aim at improving the accuracy of the reputation system, either by reducing fraud or providing adequate information. In contrast our paper shows an excess of precision in reputation may prevent entry of new users. Spagnolo (2012) and Butler et al.(2017) show in lab experiments that a reputation system, if not designed wisely may hinder entry of new participants. We provide a formal analysis of this phenomena and identify this effect in a natural experiment. Kovbasyuk and Spagnolo (2017) also show a repeated game with limited records maximizes the number of trades. This effect motivated by the fact agents may have randomly changing types. In our model, this effect is due to a trade-off between revealing information, and allowing the entry of new agents.

Finally, the motivation of the paper stems from growing literature on statistical discrimination in online markets: Edelman and Luca (2014) and Edelman et al (2016) study discrimination on the short-term house rental platform Airbnb.Goddard et al. (2015) and Ge et al. (2016) show evidence of discrimination in transportation systems. Close to our empirical analysis, Farajallah et al. (2016) shows minorities have a lower success rate on the carpooling platform Blablacar. Our paper also identifies discrimination. However, it goes into more details by showing that this discrimination is at least in part statistical. A wisely designed reputation system allowing to reveal the true quality of agents may alleviate the issue.

2 Theory models

2.1 Model I: A reputation system to alleviate the reputation trap issue

[2]In this section, we examine the impact of a reputation system on a principal's choice to interact with a given agent. The reputation, built over past interactions helps the principal figure out whether he wants to interact with an experienced agent, or exclude him and hire a new one with no reputation

[2]As abovementioned, this is a work in progress, hence we present two approaches to build a theoretical model, the first one focused more on entry under statistical discrimination and second on rationale for a reputation system on a two-sided market with moral hazard; ultimately, we aim to unify these approaches.

yet.

There are an infinity of periods, each corresponding to a ride a rider (the principal) will be doing. For each ride (each period), he needs a driver (the agent). At each period, the principal will be presented to a new agent. The principal does not observe the exact type of the agent, but has a prior on the distribution of types. For a given agent, the distribution depends on an observed attribute ϵ. The cumulative distribution function of types in a given population ϵ is denoted $F_\epsilon(.)$. At each period, the principal is matched with a new agent. The "population" from which the agent is selected is drawn randomly from CDF $G(\epsilon)$. The principal can then choose to keep the agent he was matched with at the previous period (whose population we denote ϵ_{old}), or fire her and hire the new agent. He will make his decision based on his beliefs of current and new drivers' type and propensity to exert efforts – and his own ability to elicit efforts based on the information at his disposal. The model is an extension of Laffont and Tirole (1986), applied to n periods. The principal has however no commitment power from one period to another. Hence, any information revealed and recorded by the principal, will be fully exploited at later stages of the game. For ease of exposition, all agents are myopic and disregard the impact of current actions on benefits beyond the current period. The principal aims at fully revealing, direct mechanisms. At each period i, the agent is asked to report his type $\hat{\beta}$. The agent is then paid a transfer s_i end provided with a recommendation on efforts e. Both the transfert and efforts will depend on the reported type $\hat{\beta}$, the population ϵ of the agent and the beliefs of the principal, to be described later on. The principal enjoys a gross benefit of $\pi = \beta + e_i$, meaning a high-type agent provides high benefits. The principal also wants to elicit efforts from the agent. The principal observes π, but not β and e, giving the agent the opportunity to pretend she is a low type, in order to save on efforts. The principal aims at maximizing:

$$\beta + e_i - s_i$$

The agent endures a cost of producing efforts $\psi(e)$. Her outside opportunity is 0. Her payoff if she accepts the offer of the principal is therefore:

$$u_i = s_i - \psi(e_i)$$

A key addition to standard models is that the principal only has imperfect recall of previous disclosures. Assuming agent i revealed himself to be of type β_i over the last n periods, the principal believes the type of the agent is :

$$\tilde{\beta}_{i,n+1} = (1 - \alpha^n)\beta_i + \alpha^n v_\epsilon \tag{1}$$

Where v_ϵ is randomly drawn from distribution $F_\epsilon(.)$, which corresponds to the prior distribution of types in population ϵ, absent further information. $\alpha \in [0, 1]$ is exogenous and represents how well the principal remembers the information extracted in previous periods. $\alpha = 1$ means he remembers nothing, and starts from the same prior at« every period. $\alpha = 0$ means there is perfect recall. Once the type of the agent is disclosed, he can fully exploit the information. Intermediary α means the principal has an idea of the type previously disclosed, but still relies on his prior to some extent. Another interpretation for this belief formation, is that the principal distrusts the information previously collected and still clings to his prior. We could describe this behavior as a form of persistence in prejudice. In practice α represents the strength of the reputation system, that allows to transmit private information revealed in previous stages to subsequent stages. The smaller α, the more precise the reputation system. We focus on fully revealing mechanisms. The timing within each period is:

- Step 1: principal establishes his prior on current driver (equation 1)
- Step 2: principal is matched with a new driver of population drawn from $G(\epsilon)$
- Step 3: principal chooses whether to retain his current agent, or fire her and hire the new one.
- Step 4: agent accepts/rejects participation, and reports his type $\hat{\beta}$
- Step 5: mechanism prescribes a menu of payments. Agent exerts efforts e accordingly
- Step 6: agent/principal agree on bad outcome if observed benefits do not equal $\pi(\hat{\beta}) = \hat{\beta} + e(\hat{\beta}))$

Expected benefit of a new driver The principal is matched with a driver of population ϵ. He makes sure there is full participation. We focus on direct mechanisms. Assuming the agent reports type $\hat{\beta}$, she then have to exert efforts that will replicate the total benefits $\pi(\hat{\beta})$ the principal expects to observe (otherwise the agent incurs a large penalty). Hence she has to choose

$$e(\beta, \hat{\beta}) = \pi(\hat{\beta}) - \beta$$

Her payoff if innate costs β, and reports $\hat{\beta}$ is:

$$u(\beta, \hat{\beta}) = s_i(\hat{\beta}) - \psi(\pi(\hat{\beta}) - \beta) \tag{2}$$

$$\frac{\partial U}{\partial \beta}(\beta, \hat{\beta}) = \psi'(\pi(\hat{\beta}) - \beta) \tag{3}$$

Let $U(\beta)$ be the agent's payoff with truthful reporting. The envelope theorem yields that:

$$U(\beta) = U(\underline{\beta}_\epsilon) + \int_{\underline{\beta}_\epsilon}^{\beta} \psi'(e(s))ds \tag{4}$$

To minimize rents, the principal will choose a menu such that $U(\underline{\beta}_\epsilon) = 0$. By integration by parts we get :

$$\mathbb{E}(U(\beta)) = \int_{\underline{\beta}_\epsilon}^{\bar{\beta}_\epsilon} \frac{1 - F_\epsilon(s)}{f_\epsilon(s)} \psi'(e(s)) f_\epsilon(s) ds \tag{5}$$

The expected payoff of the principal is :

$$\mathbb{E}_{new}(\Pi) = \int_{\underline{\beta}_\epsilon}^{\bar{\beta}_\epsilon} (\pi(s) - s_i(s)) f_\epsilon(s) ds \tag{6}$$

$$= \int_{\underline{\beta}_\epsilon}^{\bar{\beta}_\epsilon} \left(\beta + e - \frac{1 - F_\epsilon(s)}{f_\epsilon(s)} \psi'(e(s)) - \psi(e(s))\right) f_\epsilon(s) ds \tag{7}$$

Maximizing over the effort recommendation yields:

$$\psi'(e(\beta)) = 1 - \frac{1 - F_\epsilon(\beta)}{f_\epsilon(\beta)} \psi''(e(\beta)) \tag{8}$$

This means high types are required to exert high efforts, while low type efforts are distorted downwards. This is a classical result of the theory of incentives. Assume the distribution of types within a population ϵ corresponds to a mere shift of densities from left to right, meaning $F_\epsilon(\beta) = F_0(\beta - \epsilon)$. This means the higher ϵ , the higher the expected type of an agent. The effort recommendation then follows:

$$\psi'(e_\epsilon(\beta)) = 1 - \frac{1 - F_0(\beta - \epsilon)}{f_0(\beta - \epsilon)} \psi''(e(\beta) \tag{9}$$

Assumption 1. *For ease of exposition, we sometimes make the following assumptions:*

A1 $\psi(e) = \frac{e^2}{2}$

A2 $F_\epsilon(.)$ *follows a uniform distribution over* $[\epsilon, 1+\epsilon]$

Lemma 1. *Under assumptions 1, take an agent of a given type* β*, with no reputation. As the agent's prior signal decreases (*ϵ *small):*

1. *the required effort increases*
2. *the surplus of the agent increases*
3. *the surplus of the principal derived from the agent increases*

Proof. **Impact of ϵ on efforts:** Define :

$$h(e,\epsilon) = -\psi'(e_\epsilon(\beta)) + 1 - \frac{1-F_\epsilon(\beta)}{f_\epsilon(\beta)}\psi''(e(\beta))$$

We use the implicit function theorem on $h(e,\epsilon)$:

$$\begin{aligned}\frac{\partial e^*(\epsilon)}{\partial \epsilon} &= -\frac{1+\frac{1-F_\epsilon(\beta)}{f^2(\epsilon,\beta))}\frac{\partial f(\epsilon,\beta))}{\partial \beta}}{\psi''(e)+\frac{1-F_\epsilon(\beta)}{f(\epsilon,\beta))}\psi'''(e)} \\ &= -1 < 0\end{aligned}$$

Impact of ϵ on agent surplus: Surplus of type β from population ϵ:

$$\begin{aligned}u(\beta,\epsilon) &= \int_{\underline{\beta}_\epsilon}^{\beta} \psi'(e(s,\epsilon))f(s,\epsilon)ds = \int_{\epsilon}^{\beta}(s-\epsilon)ds = \frac{(\beta-\epsilon)^2}{2} \\ \Rightarrow \frac{\partial u(\beta,\epsilon)}{\partial \epsilon} &< 0\end{aligned}$$

Impact of ϵ on principal surplus:

$$\begin{aligned}\Pi(\beta,\epsilon) &= \beta + e(\beta,\epsilon) - \frac{1-F_\epsilon(\beta)}{f_\epsilon(\beta)}\psi'(e(\beta,\epsilon) - \psi(e(\beta,\epsilon)) \\ &= \beta+\beta-\epsilon-(1-\beta+\epsilon)(\beta-\epsilon) - \frac{(\beta-\epsilon)^2}{2} \\ \Rightarrow \frac{\partial \Pi(\beta,\epsilon)}{\partial \epsilon} &= \epsilon - \beta < 0\end{aligned}$$

□

Albeit intuitive, lemma 1 has important implications on the principal's choice of an agent. It is worth observing that absent other information than the prior $F_\epsilon(.)$, a principal will choose an agent from the highest ϵ population whenever he has the choice. Agents from a relatively low ϵ population may thus be stuck in a "*reputation trap*", whereby they get excluded from interaction – despite them having a potentially higher type than the expected type of the higher-ϵ population. This is also despite the fact they will provide more efforts if they are selected than their counterpart from a higher population –see item (1) in lemma 1– and despite the fact the agent provides more surplus to the principal –see item (3) in lemma 1. This means a reputation system is necessary to convey information from one period to the other, in order to avoid the *reputation trap* phenomena. As the next section will show, the quality of the reputation system will be key to eliminating this issue and restore efficiency.

Expected benefit of retaining current agent At the beginning the $n-th$ period with a given agent, the principal can use previous messages conveyed by the reputation system to form a prior of the agent's type. Assume the agent i is of true type β_i. The population of the incumbent is indexed by ϵ_{old}. The principal forms his beliefs according to equation (1), with $\epsilon = \epsilon_{old}$. This means the principal may doubt the report of the agent is accurate, and thinks his type will be drawn from a distribution $\tilde{F}_{\alpha,n,\epsilon}(.)$ such that $\tilde{F}_{\alpha,n,\epsilon}(x) = F_0\left(\frac{x-(1-\alpha^n)\beta_i}{\alpha^n} - \epsilon\right)$

Following similar steps as in section 2.1, we find that the expected surplus of the principal is:

$$\mathbb{E}_{old}(\Pi) = \mathbb{E}_{n,\beta_i}\left[(1-\alpha^n)\beta_i + \alpha^n v_\epsilon + e_{n,\beta_i}(s) - \frac{1-F_{\alpha,n,\epsilon}(s)}{f_{\alpha,n,\epsilon}(s)}\psi'(e_{n,\beta_i}(s)) - \psi(e_{n,\beta_i}(s))\right] \tag{10}$$

Maximizing with respect to $e_{n,\beta_i}(s)$ we find that $e_{n,\beta_i}(s)$ is defined by :

$$\psi'(e_{n,\beta_i}(s)) = 1 - \frac{1-F_{\alpha,n,\epsilon}(s)}{f_{\alpha,n,\epsilon}(s)}\psi''(e_{n,\beta_i}(s)) \tag{11}$$

While the surplus of a principal retaining its agent (10) looks similar to the one when he fires him (7), there are two key difference. First, the principal has a more precise prior on the agents' expected type (first two terms in 10). Second, the effort schedule is modified due again to a better appreciation of the agents' type. This is reflected both in the effort recommendation, and the rent left to agents.

Retention policy with a reputation system The principal will retain his agent if and only if the expected benefit from continued interaction with the current agent $\mathbb{E}_{old}(\Pi)$ is greater than the expected benefit from hiring a new agent $\mathbb{E}_{new}(\Pi)$. For ease of exposition, this section uses assumptions 1. With this assumption, we can re-write after some calculation (to be detailed in appendix A.1):

$$\mathbb{E}_{new}(\Pi) = \frac{2}{3} + \epsilon_{new} \tag{12}$$

$$\mathbb{E}_{old}(\Pi) = (1-\alpha^n)\beta_i + \alpha^n \epsilon_{old} + \frac{1}{2} + \frac{\alpha^{2n}}{6} \tag{13}$$

The current agent is retained if and only if

$$\begin{aligned} &\mathbb{E}_{old}(\Pi) > \mathbb{E}_{new}(\Pi) \\ &\Leftrightarrow \beta_i > \bar{\beta} \equiv \frac{\epsilon_{new} - \alpha^n \epsilon_{old}}{1-\alpha^n} + \frac{1}{6} + \frac{\alpha^n}{6} \end{aligned} \tag{14}$$

A first conclusion is that if α^n is close to 0 (strong reputation system), the retention policy becomes efficient, as the principal retains the manager if and only if the surplus he will obtains from continuation exceeds the expected surplus stemming from interaction with a new manager. If α^n is close to 1 (i.e. the principal has a poor memory, or the reputation system is weak), then condition 14 is met if and only if $\epsilon_{new} > \epsilon_{old}$: the principal chooses whoever has the highest public signal. In that case, agents with a low public signal are trapped in a “reputation trap". This situation is more likely to occur at early stages (small n) of the relationship between the principal and the agent (since $\frac{\partial \bar{\beta}}{\partial \alpha^n} > 0$.

If a principal with perfect recall would internalize agent surplus (in other words, if he were a benevolent social planner), he would be able to elicit optimal effort of all agents he is matched with. In that case, the retention policy is simpler:

$$\bar{\beta}_{SP} = \frac{1}{2} + \epsilon_{new} \tag{15}$$

The social planner keeps the agent if and only if its observed type is greater than the expected type of the new population. We can then write the following lemma.

Lemma 2. *Assume the principal is matched with a new agent of a higher population than the incumbent agent. There exist α_{lim} such that if $\alpha < \alpha_{lim}$, there is excess retention of the agent compared to the policy of a social planner. If $\alpha > \alpha_{lim}$, there is excess exclusion of the agent.*

Proof. There is excess retention if and only if

$$\bar{\beta} < \bar{\beta}_{SP} \Leftrightarrow w(\alpha^n) \equiv \frac{\epsilon_{new} - \alpha^n \epsilon_{old}}{1-\alpha^n} + \frac{1}{6} + \frac{\alpha^n}{6} - \frac{1}{2} - \epsilon_{new} < 0$$

$w(0) < 0$, $w(x) \underset{x \to 1}{\longrightarrow} +\infty$ and $\frac{\partial w}{\partial \alpha^n}(\alpha^n) > 0$. The lemma results from the intermediate value theorem. □

Figure 1 summarizes the findings of this section. It shows the retention policy per type β , as a function of reputation system memory α^n, when incumbent is $\epsilon_{old} = 0$ and the principal is matched with a new agent from population $\epsilon_{new} > 0$. We see that if the principal distrusts past observations or has a bad memory (α^n large), there will be excess exclusion of the incumbent agent. It may go as far as excluding all incumbent agent, notwithstanding their type. On the contrary, if the memory is too good, the principal may exert excess retention: even though the incumbent may be of relatively low type, the principal will keep her. This is because he knows her type well enough so that little rent will be given away. Closed form solutions for threshold values for α can be found in appendix A.2.

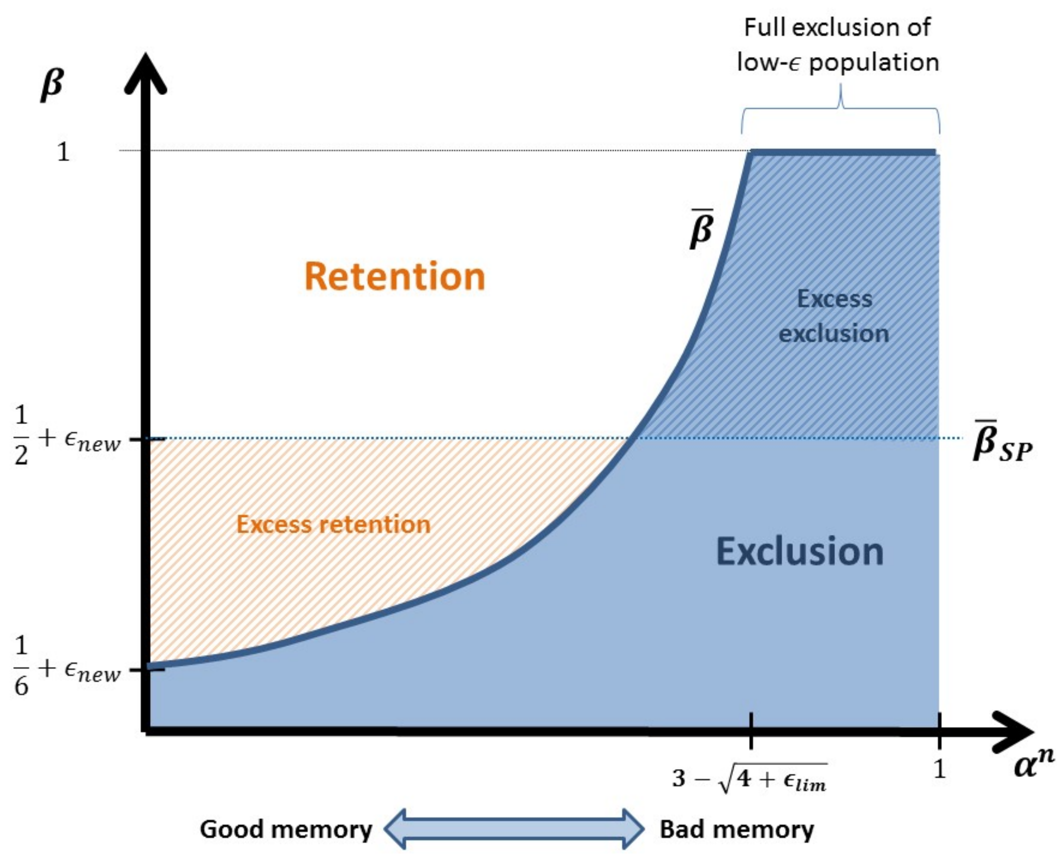

Figure 1: Retention policy per type β , as a function of reputation system memory α^n

From a social welfare perspective, it is therefore desirable that the reputation system has imperfect recall, so as to limit entrenchment of the incumbent (thanks to asymmetry of information being revealed during previous interactions) while avoiding populations to be confined in a reputation trap.

In the present model however, the principal always finds it optimal to have as good a memory as possible. A principal should therefore strive to have α as low as possible. This result is unlikely to hold if she takes into account the effect of the entry of new agents – which benefits will be reaped in future periods. The analysis of future market expansion effects on a platforms' inclination to keep a precise history of previous transactions is currently under progress.

2.2 Model II: Benchmark model of moral hazard

Consider a two-sided market populated by a monopolist platform, n_1 users on side 1 and n_2 on side 2, where $n_i \in \mathcal{N}$. We will refer to side 1 users as sellers and side 2 as buyers. Both sides of the market engage in a transaction, which is, costly for side 1 to execute, and is subject to moral hazard, i.e. there is a part of transaction that is not contractible, effects utilities of both sides of the market and depends solely on decision of side 1. Sellers are long-lived users of the platform, while buyers trade only once. Finally, there is a platform that operates the market by setting per unit transaction fees, and has a possibility to enable side 1 to build individual reputation.

Preferences and behaviour As above-mentioned, side 1 users are long-lived. If they trade they receive a benefit b, which is distributed according to some continuous distribution function $F(s)$ on the interval $[0,1]$. In each period, when trade occurs a side 1 user takes an action a, which can be either high (h) or low (l). Hence, per interaction utility of a side 1 user of type b writes 16:

$$u_i = \begin{cases} s_i - c - \tau_1 & \text{if } a = h \\ s_i + g - \tau_1 & \text{if } a = l \end{cases} \tag{16}$$

, where τ_1 is a per interaction fee set by the platform and c a cost of trading that is positive. Side 2's utility also depends on a type b_i distributed according to $H(b)$, on $[0,1]$ and on the action taken by sellers 17

$$v_j = \begin{cases} b_j - \tau_2 & \text{if } a = h \\ b_j - d - \tau_2 & \text{if } a = l \end{cases} \tag{17}$$

We assume that the immediate benefit accruing to sellers g is smaller than the damage d caused to the other side, therefore the high action is socially desirable.

Side 1 users can also be characterized by their reputation, which is observed by side 2, if a platform decides to develop a reputation system. A reputation system will allow buyers to observe previous actions of sellers, which can help them form expectations of an action that will be taken in current period. Therefore, some side 1 users might develop a reputation for always taking a high action, others for taking the low one. Consequently, side 2 users might decide with whom they want to trade. These reputational considerations are transmitted with the probability of trading $\gamma(r)$, where r is reputation level. For simplification we assume that the reputation in period t equals to the action taken in the previous period, later on we will provide conditions under which this will be true, thus: $r_t = a_{t-1}$. Therefore, the decision whether to participate or not, and which action to take is a solution to the Bellman equation 18:

$$V(s_i, r) = \gamma(r) \max\{\underbrace{0 + \delta V(s_i; r)}_{\text{do not trade}}; \underbrace{s_i + g - \tau_1 + \delta V(s_i, l)}_{\text{take a low action}}; \underbrace{s_i - \tau_1 + \delta V(s_i, h)}_{\text{take a high action}}\} + (1-\gamma)(0 + \delta V(s_i, r)) \quad (18)$$

,where $r \in \{h, l\}$. Given a strictly positive cost of trading c, trade between some users will not be desirable, there benefits will be lower than the cost, and also the platform might find it optimal to exclude some low valuation users by setting higher tariffs. Thus, for any tariffs number of users trading will be η_1 on side 1 and η_2 on side 2. Therefore, $\gamma(h) = \eta_2$, and $\gamma(l) \leq \eta_2$.

Pricing by the platform The objective of this section is to discuss the pricing problem of the platform. In a standard two-sided market set-up a platform has to weigh an incentive to increase the participation on both sides against the motivation to increase the per interaction prices. In our model, platform has to also take into account incentives of sellers to take a given action. Both the existing expectation, as this is the signal that side 2 observes, as well as expectations of future trades are the driving forces of these incentives. Full characterization of a solution to such a game corresponds to a complex dynamic game, and can be a goal in itself. For the purpose of this paper, however, we focus only on equilibria, that exhibit some interesting features.

We will restrict attention to *Stationary Markov-perfect equilibria* (MPE) in which strategies depend on the history only through payoff relevant state variables. When the platform is choosing tariffs τ, the only pay-off relevant variable at a given date is the reputation r_t, which is equal to actions taken at $t-1$, so the restriction to Markov perfection here means that the platform's strategy can be expressed as a function of reputation of its users. That is at any date t, the evolution of prices and output from date $t+1$ on depends on previous period's reputation only through the effect on current reputation. In a pure-strategy equilibrium we can write $\tau_i = \mathcal{Z}^*(r_t)$ and let $V(r_t)$ be the platform's equilibrium continuation value from period t on. When consumers are making their decisions, the state comprises both the existing reputation r_t and posted prices τ_i. Here Markov perfection requires that the active consumers behave the same way at any two histories for which their future flow of payoffs is the same under any sequence of future actions, so that the behaviour of consumers at date t is determined by τ_i and r_t.

Solution concept is Stationary Markov Perfect Equilibrium and the timing goes as follows:

1. Platform sets prices $\tau_i \in \mathbf{R}^+$ to both sides of the market
2. New cohort of side 2 users arrives, and each buyers is matched with one seller[3]; they decide whether to participate
3. Side 1 users makes participation and action decisions

Finally, we constrain platform to set one per-transaction price per side of the market. This assumption well describes the practice of many major platforms.

No reputation system We first consider the outcome of the game when the platform chooses not to have a reputation system. In order to simplify exposition, we will restrict the space of type. Consider there are two types on each side of the market, high and low; s^h and s^l, such that $s^h > s^l$. Furthermore, distribution of the types is degenerate, with probability λ_1 seller will be of the high type, and with probability $1-\lambda_1$ of a low type. Analogically, we have b^h and b^l. Furthermore, we assume that the trade between the low types is inefficient:

$$s^l + b^l + g - d < c$$

The immediate consequence of no reputation system is that there is no way a seller can signal something about her type to the other side of the market, and therefore benefit from a better reputation in the future. Lemma 1 gives the result:

Lemma 3. *Under no reputation system sellers never take a high action. They do not participate if $\tau_1 > s_i + g + \delta V_i(s_i)$, otherwise they participate and take a low action.*

[3] In the next section, we introduce competition between sellers

Proof. Problem of side 1 users writes :

$$V_i(s_i) = \gamma \max\{0 + \delta V_i(s_i); s_i + g - \tau_1 + \delta V_i(s_i); s_i - \tau_1 + \delta V_i(s_i)\} + (1-\gamma)(0 + \delta V_i(s_i))$$

We can observe that:

$$s_i + g - \tau_1 + \delta V_i(s_i) > s_i - \tau_1 + \delta V_i(s_i) \ \forall \theta$$

,because $g > 0$; if $\tau_1 > s_i + g + \delta V_i(s_i)$ their payoff is negative so they decide to not participate, for $\tau_1 \leq s_1 + g + \delta V_i(s_i)$ they participate and take the low action. □

An important implication of Lemma 1 is that side 2 users expect the low action to be taken so they will take a damage d. Therefore, only buyers with type high enough will participate i.e. $b_i - d - \tau_2 \geq 0$. Solving by backward induction we see that, first, users will participate as long as their expected benefit is higher than the prices asked by the platform, and secondly, given that the platform moves first, she has no incentives to leave them strictly better off. Therefore, prices will match the expected utility of the lowest type that the platform wants to include.

Proposition 1. *Following can be MPEs of the game:*

- *Platform charges $\bar{\tau_1} = s^h + g$ and $\bar{\tau_2} = b^h - d$; high type users on both sides of the market participate. Platform makes the per turn profit of: $\pi_t = \lambda_1 n_1 \lambda_2 n_2(\bar{\tau_2} + \bar{\tau_1} - c)$*
- *Platform charges $\underline{\tau_1} = s^l + g$ and $\bar{\tau_2} = b^h - d$; high type users of side 2 participate and both types on side 1. Platform makes the per turn profit of: $\pi_t = \lambda_2 n_2(\bar{\tau_2} + \underline{\tau_1} - c)$*
- *Platform charges $\bar{\tau_1} = s^h + g$ and $\underline{\tau_2} = b^l - d$; both types of side 2 participate and high type on side 1. Platform makes the per turn profit of: $\pi_t = \lambda_1 n_1(\bar{\tau_1} + \underline{\tau_2} - c)$*

Which of the equilibria will be reached depends on relative benefits of low and high types on both sides of the platform and on number of them.

Proof. Follows from Lemma 1, and from the observation that $\lambda_i \in [0, 1]$, $b^h > b^l$, $s^h > s^l$ so there exist parameter sets for which each of the profits can be the highest one. □

Under no reputation system users always fall for moral hazard problem, platform has no levels of linear prices to induce agents to take the high action. This is because, side 1 users are the last to take the decision on which action to choose, so they always focus on immediate benefit. The platform's only decision is whether to include only high types, or to allow low type from one of the sides. Allowing both types of both sides to participate would result in a profit loss, due to the assumption:$s^l + b^l + g - d < c$ The extent of the moral hazard issue can be seen as the gain of side 1 compared to the loss of side 2 i.e. $g - d$, which is always negative by assumption. Hence, platform has an incentive to limit amount of trade to only those agents, whose benefits are high compared to the loss caused by the moral hazard. In the equilibria from Proposition 1 this moral hazard component can be summarized in the following way: $\lambda_1 n_1 \lambda_2 n_2 (g - d)$, $\lambda_2 n_2 (g - d)$, and $\lambda_1 n_1 (g - d)$ for each of potential MPEs. The more serious the problem is, the more likely will be the platform to engage in the strategy 1, that excludes all the low types.

Model with a reputation system When a platform decides to introduce a reputation system, strategies might become more complicated. Such a system serves as an information transmission mechanism, that will provide future cohorts of side 2 with part of the history of actions of the agent on side 1. Such an information allows buyers to decide whether to trade or not with a given agent, base on her reputation. Some insights into the MPE outcomes can be obtained by comparing platform's revenue stream across steady states, where users are assured to have reputations consistent with their utility maximizing actions.

A useful way to organize discussion about potential equilibria is to divide them into pooling and separating. To a pooling equilibrium we will refer when both types of sellers maintain the same reputation, either high or low. In a separating equilibrium there will be high type users with high reputation and low type users with a low one.

For a side 1 agent to be willing to develop a high reputation the added value through higher participation $\gamma(h)$, must overcome the immediate loss of g. Individual incentive compatibility condition can be expressed as:

$$s_i - \tau_1 + \delta V_i(s_i, h) \geq s_i + g - \tau_1 + \delta V_i(s_i, l)$$

,which is equivalent to condition:

$$\delta(V_i(s_i, h) - V(s_i, l)) \geq g \tag{19}$$

Pooling equilbria: Sellers can either pool on high or on low reputation. Without restricting beliefs pooling on a high reputation cannot constitute an equilibrium because, there is no reward for giving up the immediate benefit g. Hence, each seller after being matched with a buyer who accepted to contract deviates to taking the low action. However, pooling on a low reputation can be an equilibrium:

Corollary 1. *In an equilibrium with pooling on a low reputation, a reputation system is ineffective. Proposition 1 extends to this case.*

Pooling on a low reputation is equivalent to a situation when the platform decides not to have a reputation system. Hence, if there would be any cost associated with establishing or running the reputation system, platform would not have one that pools all the types no a low reputation.

Separating equilibria: From the discussion of the pooling equilibria of the game we may conclude that the reason for the platform to have a reputation system, is to separate user based on level of their reputation, which according to the logic presented here will reflect their benefits from using the platform.

For a separating equilibrium to be reached the benefit from increased participation related to having a high reputation has to make up for a foregone immediate benefit accruing to the low action only for the high type of the seller, that is:

$$s^h - \tau_1 + \delta V_i(s^h, h) \geq s^h + g - \tau_1 + \delta V_i(s^h, l)$$

, however, for the low type the possibility of an immediate gain has to prevail:

$$s^l + g - \tau_1 + \delta V_i(s^l, l) \geq s^l - \tau_1 + \delta V_i(s^l, h)$$

In order to induce separation the platform will reward the high reputation sellers with trading with both type of buyers. Low reputation sellers will trade only with one type of buyers; platform's profit maximizing choice is to set low type sellers with only high type buyers. Proposition 2 formalizes this reasoning.

Proposition 2. *There exist a separating MPE, where platform sets prices* $\{\tau_2 = \min\left[b^l; b^h - d\right]; \tau_1 = \min[s^l + g; s^h]\}$. *That leads to per turn profits for the platform of:* $\pi_t = (\lambda_1 n_1 + (1 - \lambda_1) n_1 \lambda_2 n_2)(\tau_1 + \tau_2 - c)$. *Low type side 1 users take the low action and trade with high type on side 2. High type users on side 1 take the high action and trade with everyone on the side 2.*

Proof. Inequality 19 if satisfied for high type and not for the low type implies the solution to the 18 is:

$$V(s^h, h) = \frac{s^h - \tau_1}{1-\delta}$$
$$V(s^l, l) = \frac{\lambda_2 n_2 (s^l + g - \tau_1)}{1-\delta}$$

Incentive constraints that ensure such a separating equilibrium are:

$$V(s^h, h) \geq V(s^h, l) \iff s^h \geq \frac{\lambda_2}{1-\lambda_2} g + \tau_1$$
$$V(s^l, l) \geq V(s^l, h) \iff \frac{\lambda_2}{1-\lambda_2} g + \tau_1 \geq s^l$$

One-shot deviation principle is used to provide conditions under which agents will not deviate from their strategies:

$$V(s^h, h) \geq s^h + g - \tau_1 + \lambda_2 n_2 \delta (s^h - \tau_1) + \delta^2 V(s^h, h) \iff s^h \geq \frac{g}{(1-\lambda_2 n_2)\delta} - \tau_1$$
$$V(s^l, l) \geq \lambda_2 (s^l - \tau_1) + \delta (s^l + g - \tau_1) + \delta^2 V(s^l, l) \iff g\left(\frac{\lambda_2}{(1-\lambda_2)\delta} - 1\right) + \tau_1 \geq s^l$$

□

Reputation system provides a stick-and-carrot mechanism, sellers taking a high action are rewarded with a high participation. An increase in the amount of trade is more valuable for the high benefit type. Finally, platform sets tariffs so that it is optimal for users to separate.

Depending on the parameters, the platform can benefit from the possibility of using the reputation system. Suppose the platform charges $\underline{\tau_1}$ and $\bar{\tau_2}$, $b^h - d < b^l$ and $s^l + g > s^h$. Then, the platform's profits are higher with the reputation system because $d > g$. On top of this, there is also higher participation: $\lambda_1 n_1 + \lambda_2 n_2 - \lambda_1 n_1 \lambda_2 n_2 \geq \lambda_i \forall i$.

Importantly in the pooling equilibrium all Markov perfect equilibria are the unique equilibria of the game and hence strategies of the players do not depend on their state variables. Conversely, in the case of the separating equilibrium actions of the players depend on their current reputation, hence narrowing to this class of solutions is useful in full characterization of an equilibrium.

Again it is instructive to study how a choice of platform's strategy depends on the extend of the moral hazard problem. Compared to the pooling equilibria, moral hazard problem has been greatly resolved. Intermediary's profits do not depend on d, and increase in g, when $s^l + g > s^h$. Therefore, the more severe moral hazard problem the more likely platform is to engage in the separating equilibrium.

Lemma 4. *There exist $\bar{d}$ such that $\forall d > \bar{d}$ there is separating MPE.*

Proof. Follows directly from comparing platform's profits in Proposition 1 and Proposition 2. □

Platform decides whether to run a reputation system or not depending on multitude of factors, salience of a moral hazard issue among them, this section has introduced some basic logic, why a reputation system, might allow some users to commit to socially desirable actions, and how a platform can organize pricing in order to reward these users.

Model with competition Benchmark model presented in the previous section explains why a platform might decide to create a reputation system. However, it rests on a number of simplifying assumption, which in a consequence does not allow us to model competition between sellers, which is an important aspect of many popular two-sided markets exhibiting moral hazard e.g. Blablacar or Airbnb. Aim of this section is to introduce competition into the model. We will, firstly, allow sellers to set final prices, and secondly, buyers will be presented with a list of sellers, and they will choose based on price and reputation; finally, we drop the assumption of degenerate distribution. Sellers' benefit will be distributed according to cdf $F(s)$ on the interval $[0,1]$, and buyers's $b \sim H(b) \in [0,1]$. Finally, we want to capture a situation in which sellers compete for buyers, hence, we assume that there are N, buyers and M sellers, such that $M > N$. Timing of the game:

1. Platform sets per interaction tariffs to both sides of the market: τ_1 and τ_2
2. Side 1 users observe tariffs τ_1 and τ_2, decide whether to participate and set final price $p_i \in [0,1]$
3. Side 2 users arrive make a query and, are matched with a set of users N, which is a natural number, they decide with whom, if any to trade.
4. Side 1 users take action, payoffs are realized

Solution concept is Stationary Markov Equilibria, in which state variable will be reputation of side 1 users. We solve the game by backward induction.

Pooling equilibria: Firstly, we will discuss outcomes in which all users take the low action. This could be because the platform has not developed a reputation system or it has set tariffs that induced such a result. In this case sellers always take the low action. Thus, the payoff of a buyer j trading with a seller i writes:

$$v_j = b_j - d - \tau_2 - p_i^*$$

, where p_i^* is the optimal price set by the seller i. Problem of a buyer is, having observed the prices, to decide whether to trade and with which seller. Therefore, problem of the buyer j is:

$$\max \{b_j - \tau_2 - p_1(s_1), b_j - \tau_2 - p_2(s_2), ..., b_j - \tau_2 - p_N(s_N), 0\}$$

, which boils down to choosing the seller with the lowest price, or not trading. Given a price $p_i(s_i)$ probability of a trade occurring is $Pr(b_j \geq \tau_2 + g + p_i(s_i)) = 1 - H(b_j)$

Tariffs are set in a way that it is optimal for sellers to develop a low reputation i.e.: $V(s_i, l) \geq V(s_i, h) \forall s_i$. Sellers know the number of competitors, but only their own valuation. They set prices independently and the seller setting the lower price gets selected. This set-up resembles the first prices auction, hence following the *Revelation Principle* user with the highest benefit will set the lowest price. However, the price has also to satisfy $b_j \geq \tau_2 + g + p_i(s_i$, so that a buyer will decide to trade. Utility function is described by:

$$u_i = (s^i + p_i - c - \tau_1) Pr(p_i \leq \min p_j \forall j \notin i \ \& \ P_i \leq b_i - d - \tau_2)$$

which given symmetrical strategies played by sellers is equivalent to:

$$u_i = (s^i + p_i - c - \tau_1) Pr(s_i \geq \max s_j \forall j \notin i \ \& \ P_i \leq b_i - d - \tau_2)$$

The range of $p(\cdot)$ in an interval $p([0,1]) = [\underline{p}; \bar{p}]$; thus without loss of generality we may suppose that $p_i \in [\underline{p}; \bar{p}]$, so there exists $x \in [0,1]$ such that $p(x) = p_i$, therefore the problem of a side 1 users is equivalent to solving following maximization problem:

$$\max_{p_i} (\hat{s}_i + p_i(x)) F(x)^{N-1} (1 - H(p_i(x)) \tag{20}$$

,where $F(\cdot)$ is a cdf of a distribution of types on side 1, $f(\cdot)$ will be its pdf, and $\hat{s}_i = s^i - c - \tau_1 + g$. Lemma provides a solution.

Lemma 5. *Price set by a seller i with valuation s_i is a solution to a following differential equation:*

$$p(s_i) = s^i - c - \tau_1 + g + \frac{\int_0^s F(u)^{N-1}(1-G(p(u)))du}{F(s)^{N-1}(1-G(p(s)))} \tag{21}$$

Proof. Taking first order condition of 20 we obtain:

$$\hat{s}_i \left[(N-1)f(s_i)F^{N-2}(s_1)(1-G(p(s_i))) - F^{N-1}g(p(s_i))p'(s_i)\right] \tag{22}$$

$$= p'(s_i)F^{N-1}(s_i)(1-G(p(s_i))) + f(s_i)F^{N-2}(s_i)p(s_i)(1-G(p(s_i)))(N-1) - p(s_i)F^{N-1}(s)g(p(s_i))p'(s_i) \tag{23}$$

Note, that:

$$(F(s)^{N-1}(1-G(p(s_i)))' = (N-1)f(s_i)F^{N-2}(s_1)(1-G(p(s_i))) - F^{N-1}g$$

$$(p(s)F(s)^{N-1}(1-G(p(s))))' = p'(s_i)F^{N-1}(s_i)(1-G(p(s_i))) + f(s_i)F^{N-2}(s_i)p(s_i)(1-G(p(s_i)))(N-1) - p(s_i)F^{N-1}(s)g$$

Combining 23, 24 and 25 we have:

$$\hat{s}_i \left[F(s_i)^{N-1}(1-G(p(s_i)))\right]' = \left[p(s_i)F(s_i)^{N-1}(1-G(p(s_i)))\right]'$$

Integrating from 0 to s, we have:

$$\int_0^s \hat{s}_i \left[F(s_i)^{N-1}(1-G(p(s_i)))\right]' ds = \int_0^s \left[p(s_i)F(s_i)^{N-1}(1-G(p(s_i)))\right]' ds$$

integrating by parts and observing that if $s \to 0$, than $p(s) \to 0$ because $p(\cdot)$ is bounded so $k = 0$, we obtain optimal price:

$$p^*(s_i) = \begin{cases} s^i - c - \tau_1 + g + \frac{\int_0^s F(u)^{N-1}(1-G(p(u)))du}{F(s)^{N-1}(1-G(p(s)))} \text{ if } s \neq 0 \\ 0 \text{ if } s = 0 \end{cases} \tag{26}$$

□

We can observe that the optimal price differs from the standard first price auction, because of the two-sidedness of the market. Side 1 users have to take into account participation on the other side of the market. Increasing price reduces probability of trade, so sellers have to strike a right balance, between increasing per transaction profit and likelihood of one occurring. Its hard to provide a general solution to 21, however, we can provide some insights. Equilibrium prices posted by users are functions of their benefits from trading, tariffs set the platform, distribution of benefit of buyers and importantly of the number of competitors. The higher the benefit, the lower the price and the more competitors the lower the price. Note, that because of the impact on probability of trading, sellers cannot fully pass through increases in τ_1; this is important because shows that despite the fact that sellers set prices, platform will be still able to make some profit on them.

Finally, platform sets the tariffs for both sides of the market in order to maximize net present values of its profits, subject to i. In a Stationary Markov Perfect Equilibrium profits in each period will be the same, thus platform's problem simplifies to both types preferring taking the low action:

$$\max_{\tau_1,\tau_2} \pi = \max_{\tau_1,\tau_2} \left\{(\tau_1+\tau_2)(1-F(\underline{s})(1-H(\underline{b})^N\right\} \quad (27)$$
$$\text{subject to: } V(s_i,l) \geq V(s_i,h) \forall s_i$$

, where $\underline{b} = \tau_2 + d + p^*(1)$ and $\underline{s} = \tau_1 + c - g - p^*(s_{\underline{s}})$ describe the lowest type of each side number of users trading on each side of the market. There is N buyers and each of them chooses a seller to trade, hence the product of two cdfs is the total amount of trade.

3 Empirics of moral hazard on a platform

Aim of this section is to try to validate some of the intuitions developed in theoretical part. Firstly, we are interested in investigating the value of reputation for users. Secondly, we want to see whether moral hazard can be a problem on a platform and whether high reputation can be a way to escape it. Major platforms typically restrict access to their data. Therefore, we have collected our dataset using a web-crawler application on a BlaBlaCar.fr website from 1.07.2017 to 10.08.2017. Furthermore, we have matched collected data with several other datasets: origins of names are matched with French government index [4], value of cars is matched with a average price of a the same type of car posted on ebay in Germany, distances and expected time in public transportation are calculated using google.maps, suggested and maximum prices are calculated following a cost-based formula of BlaBlaCar, and index of crime is matched with French government statistics. We have selected routes starting and finishing in radii of one of 400 largest French cities *not all data is yet collected.*

In these preliminary empirical results, we define minority agents as these whose name has an

[4] https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/

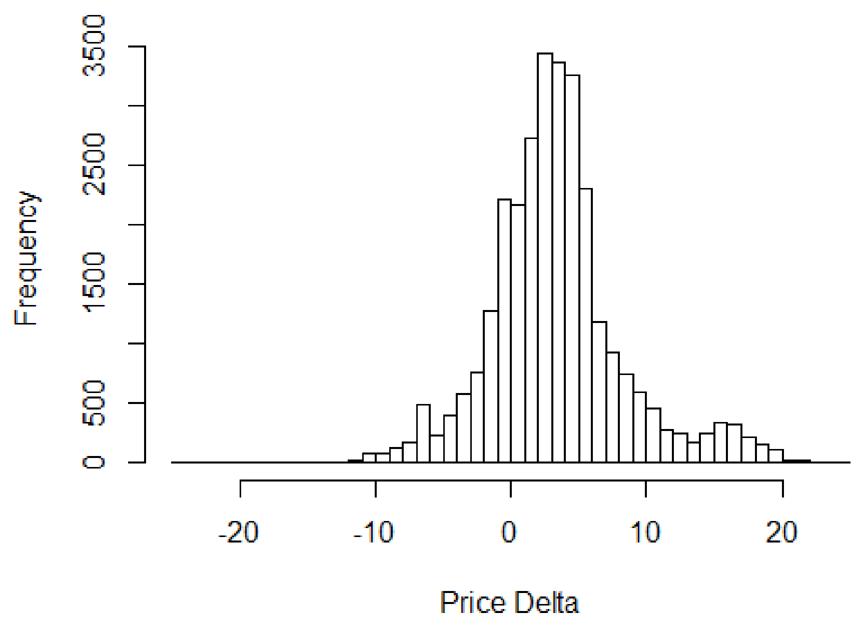

Figure 2: Histogram Price Delta

Arabic origin. While we aim at extending this definition to other origins, we motivate this definition by previous research showing that there is meaningfully lower demand for drivers with Arabic-sounding names, despite them setting significantly lower prices (see Farajallah et al (2016)).

3.1 Overview of data

On BlablaCar website we have drivers, who have empty places in their cars, interact with people looking for a ride. Drivers' strategic decisions is, foremost, setting a price and number of places to offer. BlaBlaCar claims that its mission is to help drivers recoup some of the costs related to driving the car, rather than run a professional transportation endeavour. This, apart from being an attractive marketing slogan, results also in BlablaCar suggesting prices to drivers, and setting a maximum price. Furthermore, passengers are informed whether a price offered is a deviation (up or down) from a suggested price. Thus, decision of drivers is really whether and how much to deviate from the suggested price, we call this variable price delta and figure 1 shows it.
histogram. Distribution of price delta resembles the normal one, there is small skewness to the right, but the vast majority of drivers set prices within 5 EUR range around 0. The suggested price is based

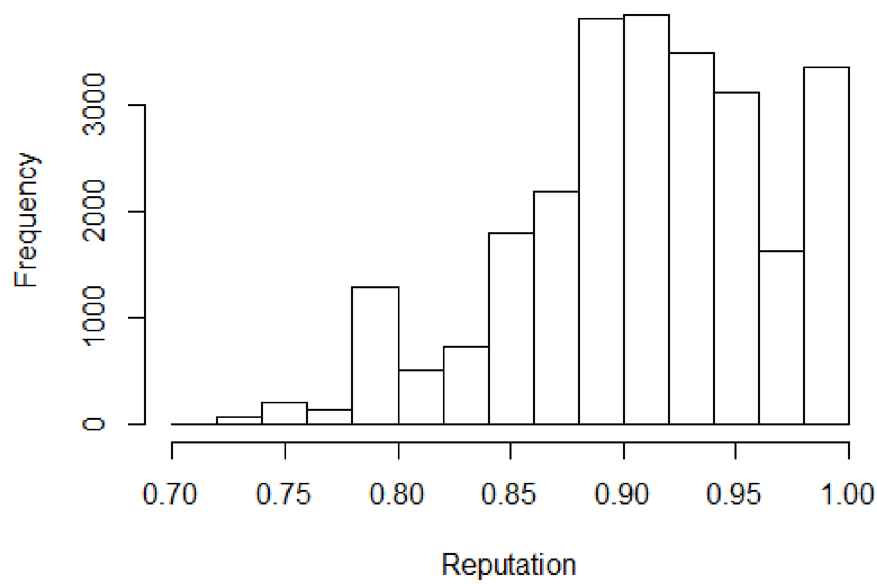

Figure 3: Histogram Average Reputation

on expected fuel consumption for an average car and the distance to be covered. Therefore, deviations can be to some extend explained by drivers having cars with higher/lower average fuel consumption. Other than cost elements, factors such that competition from other drivers on the Blablacar as well as other transportation options can have influence on pricing. Finally, and most importantly from our perspective a moral hazard related strategic issues, someone having low reputation might be inclined into pricing lower, as well as someone driving from a dangerous neighbourhood.

Characteristic feature of many online reputation systems is that most users leave very high reviews, and in result there is little variation in data. This is the case also with BlaBlaCar. Drivers are evaluated on a scale from 0 to 5, and there is possibility of leaving a written commentary. We have divided average grades by 5, so that they span from 0 to 5, and average grade is 0.76. Figure 2 is a histogram.

After collecting and matching the data we have deleted outliers and we are left with approximately 16.000 observations. Av average price of a ride is 20 euros and it is for travelling a distance of 250 km in a car worth 13.00 euros. Average driver is 36 years old, is using the platform for about a year. Most of drivers are men (65%) and a little more than half have a name of a French origin. Table 1 provides a summary statistics.

Statistic	N	Mean	St. Dev.	Min	Max
ride price	30,425	26.135	16.670	2	218
driver age	30,520	37.419	12.934	18	101
number reviews	26,403	46.256	81.623	1	1,482
number rides	26,244	36.376	42.942	2	206
Arabic	24,676	0.075	0.264	0	1
French	24,676	0.691	0.462	0	1
reputation	26,129	0.918	0.057	0.740	1.000
seniority months	30,522	39.127	27.843	0	152
male	24,676	0.693	0.461	0	1
car price	25,833	5,738.693	4,487.186	832.982	22,627.740
km	30,522	347.478	229.575	0.000	944.674
duration public transport	26,784	187.029	126.157	24.632	877.083
suggested price	30,522	22.586	14.922	0.000	61.404
price delta	29,776	3.363	4.938	−11.581	21.324
crime origin	27,232	8.996	3.038	1.840	31.270
crime destination	25,696	8.531	3.139	1.840	31.270
empty seats	30,522	2.365	0.889	0	4
taken seats	30,522	0.335	0.653	0	4

Table 1: Summary Statistics

3.2 Simple linear regressions – cross-section

A hypothesis stemming from the theory model presented in earlier sections is that salience of moral hazard matters for profits. Individuals who are perceived to either be more likely to take a low action or whose low action lead to higher damage, should expect to have lower payoffs, unless they proof to be of a high type. Therefore, these users have higher incentives to form a reputation of taking the high action. In the environment of Blablacar this means charging low prices to increase probability of a match, which means that they have an opportunity to build reputation.

We focus on two variables that can lead to a belief that someone is more likely to be of a low type: male and with Arabic name. We show several linear regressions where the explained variable is price delta, and explanatory variables are price of the car, which is a proxy of quality, duration of public transport that measures attractiveness of the outside option, number of offers made before by the driver that measures experience, male, Arabic name, and a product of male and Arabic name. Furthermore, we run our model on three samples; firstly, entire dataset, secondly, on a group of drivers that have made less than 10 offers before, so they are newcomers to the platform and don't have a well established reputation, and finally on these users that have made more than 25 offers, so these are experienced users with a formed reputation. We expect that moral hazard parameters will matter more when users

	Entire sample	Newcomers	Newcomers	Experienced	Experienced
(Intercept)	0.57	−0.94	−0.87	0.44	0.26
	(0.66)	(0.95)	(0.95)	(1.21)	(1.21)
reputation	2.64***	4.21***	4.23***	2.05	2.27
	(0.70)	(0.97)	(0.97)	(1.31)	(1.31)
car price	0.00	−0.00	−0.00	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
duration public transport	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
number reviews	−0.00***	0.03	0.03	−0.00***	−0.00***
	(0.00)	(0.03)	(0.03)	(0.00)	(0.00)
male	−0.31***	−0.18	−0.35*	−0.15	−0.21
	(0.09)	(0.15)	(0.15)	(0.13)	(0.12)
Arabic	0.42	2.86***		0.03	
	(0.37)	(0.59)		(0.62)	
male*Arabic	−0.97*	−3.49***		−0.65	
	(0.41)	(0.67)		(0.65)	
R^2	0.02	0.02	0.02	0.02	0.01
Adj. R^2	0.02	0.02	0.02	0.01	0.01
Num. obs.	15967	5222	5222	6761	6761
RMSE	4.85	4.97	4.98	4.44	4.45

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: OLS models

do not have much of reputation (columns 2 and 3), and once reputation is formed they don't matter anymore. Table 2 present the results.

We see that indeed when users do not have reputation, moral hazard matters, men and especially minority males set lower prices at the beginning of their career on Blablacar platform. After users have formed their reputations, and the types have been revealed it moral hazard parameters don't matter, reputation is more important.

3.3 Simple linear regressions – analysis of review trends

A prediction of our model is that drivers suspected of being low type, while being actually high type will exert a lot of efforts at the beginning of their career with Blablacar, in order to build reputation and escape the 'low-reputation' trap. Only once reputation is established and the issue of asymmetry of information has been alleviated, can the driver ease off on efforts. In this case, we can expect these drivers to be awarded particularly high grades during their first rides, and then converge to slightly lower grades, in line with other drivers.

To observe this, we first calculate for each driver, the trend of grades he received over his career. A decreasing trend is interpreted as a sign of high effort at the beginning of the career and then some easing. This trend is used as a dependent variable in table (3). The trend is generally decreasing, especially for people with an Arabic-sounding name. This happens despite them receiving grades similar to those received by the rest of the sample, when one restricts one's attention to later stages of the career [this later point has not been proven yet].

Table 3: Regression Results – trend in grades received

	Dependent variable:
	grade_trend/5
driver age	0.00000
	(0.00001)
male1	0.0001
	(0.0002)
Arabic	−0.002***
	(0.001)
car price)	0.000
	(0.00000)
length bio	0.00000
	(0.00000)
length ride	0.00000
	(0.00000)
km	−0.00002**
	(0.00001)
number reviews	0.00001***
	(0.00000)
driver skill	0.001***
	(0.0001)
Constant	−0.008***
	(0.001)
Observations	1,959
R^2	0.071
Adjusted R^2	0.066
Note:	*p<0.1; **p<0.05; ***p<0.01

To build reputation, populations suspected of being a low type need to escape the 'low-reputation trap". They can do so by turning preferably to a population less inclined to perceive them as low type. To observe this behavior, one can take advantage of the history of grades received by the driver. Focusing on the origin of riders' names, we observe that (1) people whose name indicates they are from a minority tend to ride with people also with a minority name and (2) this is less true as the number of rides for a given driver increases. We interpret this pattern as follows: minority drivers may escape the low-reputation trap by accumulating good grades from a population less inclined to perceive them as low-type at the beginning of their career. Then, once reputation is established the origin of the name no longer matters (see previous section) and drivers can mingle with the whole set of riders. The reputation system therefore acts as a powerful tool to mitigate statistical discrimination.

To identify this, we first run a regression of whether the riders taken onboard have a minority name or not. This regression is done individually for each driver. The intercept and trend will be used as dependent variables in table 4. The intercept (column 2) reflects the origin of riders a driver starts with, while the trend (column 3) describes whether the average origin of riders varies over time.

Table 4: Regression Results – Intercept and trend of carpoolers with arabic names

	Dependent variable:	
	name_intercept	name_trend
	(1)	(2)
driver age	0.0002**	−0.00000
	(0.0001)	(0.00001)
male	0.003	−0.00002
	(0.003)	(0.0003)
Arabic	0.025***	−0.004***
	(0.008)	(0.001)
car price	−0.00000	0.000
	(0.00000)	(0.000)
length bio	0.0001	0.00000
	(0.0001)	(0.00001)
length ride	−0.00002	0.00000
	(0.00002)	(0.00000)
km	−0.0001	−0.00001
	(0.0001)	(0.00001)
number reviews	−0.00001	0.00001***
	(0.00001)	(0.00000)
driver skill	0.002	−0.0002
	(0.002)	(0.0002)
Constant	0.006	−0.001
	(0.017)	(0.002)
Observations	2,859	2,735
R^2	0.008	0.032
Adjusted R^2	0.005	0.028
Note:	*p<0.1; **p<0.05; ***p<0.01	

4 Reputation system as a Preemptive device

Aim of this section is to identify circumstances under which incumbent platform would change its strategy to preempt entry. In particular, to discuss how a barrier to entry can be formed by using a separating strategy.

Suppose that at the time t, arrives a more efficient entrant, that has to pay entry cost f, if decides to enter. By trading with an entrant agents receive additional benefits μ_1 and μ_2 that correspond to the better technology, eq 28 describes the per period profits:

$$u_{t,1} = \begin{cases} b_1^\theta + \mu_1 - \tau_1 & \text{if } a = h \\ b_1^\theta + g + \mu_1 - \tau_1 & \text{if } a = l \end{cases} \quad u_{t,2} = \begin{cases} b_2^\theta + \mu_2 - \tau_2 & \text{if } a = h \\ b_2^\theta - d + \mu_2 - \tau_2 & \text{if } a = l \end{cases} \tag{28}$$

We consider that after entry of a more efficient player the (previous) incumbent remains a competitive fringe, that means she is ready to launch her production, had the entrant started charging too high prices. This results in several potential entry strategies of a more efficient firm[5]; she can try taking the whole market, or focusing on one of the types, finally if separating equilibrium is the one granting the highest profits, entry becomes more difficult, because earned reputation is platform specific, therefore high type side 1 users may face a question whether they are willing to lose their reputation, which gives them a higher participation rate for an exchange of higher benefit μ_1, the incumbent can exploit this dilemma by inducing separation, and thus foreclosing the market. This section aims at showing this mechanism. Timing of the game is adjusted, to capture the decision of the entrant who observes strategy of the incumbent and decides to enter or not:

1. Incumbent sets prices
2. Entrant observes prices and reputations, decides whether to enter or not. If she enters, she pay a fixed cost f and sets own prices τ_i^e.
3. Side 1 users observe prices and decide on which platform to trade
4. New cohort of side 2 arrives, observes prices, participation decisions of side 1 and reputation if side 1 users remained on the incumbent. They decide on which platform to trade. They are matched with a side 1 users and decide to participate or not.
5. Side 1 users decide to participate or not, and which action to take.

[5] We should explore an idea of platforms co-existing

We will assume that the technological advantage of an entrant is small enough, so that the benefit from trade between the low types when the low action is taken is still lower than the cost of executing this operation: $b_1^l + b_2^l + g - d + \mu_1 + \mu_2 < c$

Lemma 6. *In a* ***pooling equilibrium*** *entrant charges* $\tau_1^e + \tau_2^e = \mu_1 + \mu_2 + c$*. Suppose, all side 1 users trade and* λ_2 *of side 2 users. Entrant makes per turn profit of* $\pi_t = \mu_1 + \lambda_2\mu_2$*, and she enters the market if* $\Pi = \sum_{t=1}^{\infty} \pi_t = \frac{\mu_1 + \lambda_2\mu_2}{1-\delta} \geq f$

Proof. To be added formally. Intuitively Bertrand competition brings prices to marginal costs; side 1 users make their participation decision first, that decision will be observed in the next stage by side 2, therefore users will coordinate where side 1 decides to trade. Hence, competitors have to guarantee that side 1 users are on board, and they will charge the side 2 accordingly that is, prices will satisfy the following:

$$b_1^l + g + \mu_1 - \tau_1 \geq 0$$

$$b_2^h - d + \mu_2 - \tau_2 \geq 0$$

$$\tau_1 + \tau_2 = c + \mu_1 + \mu_2$$

□

Pooling equilibrium leads to a direct competition between the players, incumbent fails to attract any consumers and entrant is only able to charge her technological benefit on both sides. Finally, $1 - \lambda_2$ side 2 users are left out. Consumers benefit from this competition, because of lower prices.

In a **separating equilibrium** incumbent provides an additional benefit to its installed customers, she gives side 2 users an additional advantage of being able to distinguish between high and low type side 1 users. This is their payoff is: $u_2^l = \lambda_1 \left[b_2^l - \tau_2\right]$, when they observe that they have been matched with low type side 1 user, they expect a low action and opt out. On the other side, while trading with entrant they receive: $u_2^l = b_2^l - (1-\lambda_2)d - \tau_2^d$, there is a probability $1 - \lambda_2$ of being matched with the low type user who will take the low action, and furthermore they have to pay for sure.

Let's define as $\mathcal{S}_p$ the total surplus created by trade on each platform at the period of entry t, where $p \in \{i, e\}$ the incumbent or the entrant. Trade on the incumbent platform brings:

$$\mathcal{S}_I = \lambda_2 \left(b_2^h - (1-\lambda_1)d\right) + (1-\lambda_2)\lambda_1 b_2^l + \lambda_1 b_1^h + (1-\lambda_1)\lambda_2(b_1^l + g) - (\lambda_1 + \lambda_2 - \lambda_1\lambda_2)c$$

and, on entrant's platform:

$$\mathcal{S}_E = \lambda_2 b_2^h + (1-\lambda_2)(b_2^l - d) + \mu_2 + \lambda_1 b_1^h + (1-\lambda_1)(b_1^l + g) + \mu_1 - c$$

The difference between the surpluses characterizes competitive situation:

$$\Delta\mathcal{S} = \lambda_1(d + \lambda_2 c) - (1-\lambda_1)(1-\lambda_2)\left(b_2^l - b_1^l - g\right) - \mu_1 - \mu_2$$

If an entrant decides to enter and manages to attract consumers to join her platform it provides additional flow of discounted benefits of:

$$\sum_{t=1}^{\infty} \delta^t \mathcal{S}_E = \frac{\delta(\lambda_1\mu_1 + (1-\lambda_1)\lambda_2\mu_2)}{1-\delta}$$

Therefore, if the total additional surplus created by entrant, which is: $-\Delta\mathcal{S} + \sum_{t=1}^{\infty} \delta^t \mathcal{S}_E$ is higher than f, entrant will enter in the period she arrives, potentially offer negative prices in the first period if: $\Delta\mathcal{S} > 0$, but will take over the market and recap losses from the first period.

Entry to the market in which there is separation equilibrium played is additionally difficult, because an entrant has to compensate the low type of side 2 for inability to distinguish between the types with which they trade.

An interesting feature of this model is this dynamic aspect, when an incumbent learns about prospective entry to the market. There is a threshold of entry costs f, for which an incumbent who in the absence of entry induces pooling equilibrium, switches to separating one, in the attempt of entry. For this to be case, first, pooling has to be optimal and move to separation has to be effective.

Pooling will be played if:

$$\max\{\lambda_1\lambda_2\left(\bar{\tau}_2 + \bar{\tau}_1 - c\right); \lambda_2\left(\bar{\tau}_2 + \underline{\tau_1} - c\right); \lambda_1\left(\underline{\tau_2} + \bar{\tau}_1 - c\right)\} \geq (\lambda_1 + (1-\lambda_1)\lambda_2)\left(\min\{b_1^l + g; b_1^h\} + \min\{b_2^l; b_2^h - d\} - c\right) \quad (29)$$

Level of entry costs $\underline{f}$ for which there is entry depends on the equilibrium reached it can be either:

$$\frac{\mu_1 + \lambda_2\mu_2}{1-\delta} = \underline{f}$$

or:

$$\frac{\lambda_1\mu_1 + \mu_2}{1-\delta} = \underline{f}$$

The critical value of $\bar{f}$ that would prevent entry into the separating equilibrium is described by the differences in surpluses in the period of entry plus future profits, that is:

$$\bar{f} \geq -\Delta \mathcal{S} + \frac{\delta}{1-\delta}(\lambda_1 + (1-\lambda_1)\lambda_2(\tau_1^e + \tau_2^e - c) \tag{30}$$

,where $\tau_1^e = \min\{b_1^l + g; b_1^h\} + \mu_1$ and $\tau_2^e = \min\{b_2^l; b_2^h - d\} + \mu_2$.

Proposition 3. *If $f \in [\underline{f}; \bar{f}]$ upon learning that about entry incumbent changes its strategy from pooling to separating, and by doing so she forecloses the market.*

Proof. To be added. Entrant can enter through pooling or separating, so both conditions should be here. □

Proposition 3 shows that a decision, whether to develop a reputation system can be driven by competitive effects. By limiting the type space we have constrained the choice of design of reputation system, into basically the choice of having one; next section will discuss some choices of its design.

5 Conclusion

While it is well-known that a platform can use prices to steer participation the present paper shows that participation is also a function of the reputation system in place. We observe that an excessively precise reputation system may lead to the entrenchment of incumbents, and limited market expansion. Similarly a system with poor precision paves the way for statistical discrimination and little entry, especially when it comes to populations facing discrimination. Further, we show that an over-selective platform, allowing only high types to participate may be less profitable than a more lenient one that allows less-performing agents to participate. The paper describes conditions under which the platform wants to implement selection with a reputation system or not. Our empirical analysis uses unique data of transactions on a famous online carpooling platform. Similarly to previous studies it shows that populations perceived to be of a low type are discriminated against and have to set lower prices. However, this is especially true at early stages of interactions on the platform. Once reputation is built, little discrimination is observed. To build their reputation, minorities tend to lower their prices and interact mostly with agents who themselves belong to a minority.

This research highlights the importance of well-designed reputation systems in order to induce optimal entry and market expansion. Early stages of reputation building are particularly important.

Solutions include providing new members with a "default" rating set e.g. at the average rating of the market (see Spagnolo (2012)). To gain credibility, this rating would probably need to substantiated by measures of moral hazard obtained from outside the platform (social networks, rating in other online communities etc.). This however raises the issue of data privacy and ownership between online platforms and participants.

References

[1] Bar-Isaac, H. and Tadelis, S. (2008). Seller Reputation. *Foundations and Trends® in Microeconomics*, 4(4):273–351.

[2] Baron, D. P. and Myerson, R. B. (1982). Regulating a monopolist with unknown costs. *Econometrica: Journal of the Econometric Society*, 50(4):911–930.

[3] Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science*, 50(11):1587–1602.

[4] Cabral, L. L., Hortacsu, A., and Hortaçsu, A. (2010). The dynamics of seller reputation:Evidence from eBay. *The Journal of Industrial Economics*, LVIII(1):54–78.

[5] Caillaud, B. and Jullien, B. (2003). Chicken & Egg: Competition among Intermediation Service Providers. *The RAND Journal of Economics*, 34(2):309.

[6] Edelman, B. G. and Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *SSRN Electronic Journal*.

[7] Farajallah, M., Hammond, R. G., and PPnard, T. (2016). What Drives Pricing Behavior in Peer-to-Peer Markets? Evidence from the Carsharing Platform BlaBlaCar. *SSRN Electronic Journal*.

[8] Fudenberg, D. and Tirole, J. (2000). Pricing a network good to deter entry. *The Journal of Industrial Economics*, XLVIII(4):373–390.

[9] Garrett, D. F. and Pavan, A. (2012). Managerial Turnover in a Changing World. *Journal of Political Economy*, 120(5):879–925.

[10] Ge, Y., Knittel, C., MacKenzie, D., and Zoepf, S. (2016). Racial and Gender Discrimination in Transportation Network Companies. *NBER Working Paper Series*, (22776):1–38.

[11] Goddard, T., Kahn, K. B., and Adkins, A. (2015). Racial bias in driver yielding behavior at crosswalks. *Transportation Research Part F: Traffic Psychology and Behaviour*, 33:1–6.

[12] 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.

[13] Jullien, B. and Park, I. U. (2014). New, like new, or very good? Reputation and credibility. *Review of Economic Studies*, 81(4):1543–1574.

[14] Laffont, J.-J. and Tirole, J. (1986). Using Cost Observations to Regulate Firms. *Journal of Political Economy*, 94(3):614–641.

[15] Liu, Q. and Skrzypacz, A. (2014). Limited records and reputation bubbles. *Journal of Economic Theory*, 151(1):2–29.

[16] Livingston, J. A. (2005). How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions. *Review of Economics and Statistics*, 87(September):453–465.

[17] Maskin, E. and Tirole, J. (1988). A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs. *Econometrica*, 56(3):549–569.

[18] Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation.

[19] Nosko, C. and Tadelis, S. (2015). The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment. *NBER Working Paper Series*, page 20830.

[20] Rochet, J.-C. and Tirole, J. (2003). Two-Sided Markets. *Journal of the European Economic Association*, pages 990–1029.

[21] Roger, G. and Vasconcelos, L. (2014). Platform Pricing Structure and Moral Hazard. *Journal of Economics and Management Strategy*, 23(3):527–547.

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

[23] Zervas, G., Proserpio, D., and Byers, J. (2015). A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. *Where Every Stay is Above ...*, pages 1–22.

A Retention and exclusion decisions

A.1 Expected revenues from an agent

Expected revenues of a new agent: Equation 12 is derived by plugging assumption 1 in equation 10:

$$\mathbb{E}_{new}(\Pi) = \mathbb{E}_{n,B_i}\left[(1-\alpha^n)\beta_i + \alpha^n v_\epsilon + e_{n,\beta_i}(s) - \frac{1-F_{\alpha,n,\epsilon}(s)}{f_{\alpha,n,\epsilon}(s)}\psi'(e_{n,\beta_i}(s)) - \psi(e_{n,\beta_i}(s))\right]$$

Since the agent is new, we take $n = 0$ (no history):

$$= \mathbb{E}_{0,\beta_i}\left[v_\epsilon + e_{0,\beta_i}(s) - \frac{1-F_{\alpha,0,\epsilon}(s)}{f_{\alpha,0,\epsilon}(s)}e_{0,\beta_i}(s) - \frac{(e_{0,\beta_i}(s))^2}{2}\right]$$

$\mathbb{E}_{0,\beta_i}[v_\epsilon] = \frac{1}{2} + \epsilon$. Profit maximization yields 9, or $e_{0,\beta_i}(s) = s - \epsilon$

$$\begin{aligned} &= \frac{1}{2} + \epsilon + \int_\epsilon^{1+\epsilon}(s-\epsilon)\left[1 - \frac{1-(s-\epsilon)}{1} - \frac{s-\epsilon}{2}\right]ds \\ &= \frac{1}{2} + \epsilon + \frac{1}{6} = \frac{2}{3} + \epsilon \end{aligned}$$

Expected revenues of incumbent agent: Equation 13 is derived by plugging assumption 1 in equation 10:

$$\begin{aligned} \mathbb{E}_{old}(\Pi) &= \mathbb{E}_{n,B_i}\left[(1-\alpha^n)\beta_i + \alpha^n v_{\epsilon_{old}} + e_{n,\beta_i}(s) - \frac{1-F_{\alpha,n,\epsilon_{old}}(s)}{f_{\alpha,n,\epsilon_{old}}(s)}\psi'(e_{n,\beta_i}(s)) - \psi(e_{n,\beta_i}(s))\right] \\ &= (1-\alpha^n)\beta_i + \alpha^n\left(\frac{1}{2} + \epsilon_{old}\right) + \mathbb{E}_{n,\beta_i}\left[e_{n,\beta_i}(s) - \frac{1-F_{\alpha,n,\epsilon_{old}}(s)}{f_{\alpha,n,\epsilon}(s)}e_{n,\beta_i}(s) - \frac{(e_{n,\beta_i}(s))^2}{2}\right] \end{aligned}$$

Note that $F_{\alpha,n,\epsilon_{old}}(s) = \frac{x-(1-\alpha^n)\beta_i}{\alpha^n} - \epsilon_{old}$ and is defined over $[(1-\alpha^n)\beta_i + \alpha^n\epsilon_{old}, (1-\alpha^n)\beta_i + \alpha^n(1+\epsilon_{old})]$. $f_{\alpha,n,\epsilon_{old}}(s) = \frac{1}{\alpha^n}$ over the same interval, 0 otherwise. Profit maximization yields 9, or

$e_{n,\beta_i}(s) = 1 - \frac{1-F_{\alpha,n,\epsilon_{old}}(s)}{f_{\alpha,n,\epsilon_{old}}(s)} = (1-\alpha^n)(1-\beta_i) + s + \epsilon_{old}\alpha^n$. It follows:

$$\begin{aligned}
\mathbb{E}_{old}(\Pi) &= (1-\alpha^n)\beta_i + \alpha^n\left(\frac{1}{2}+\epsilon_{old}\right) + \int_{(1-\alpha^n)\beta_i+\alpha^n\epsilon_{old}}^{(1-\alpha^n)\beta_i+\alpha^n(1+\epsilon_{old})} e_{n,\beta_i}(s)\left(1-(1-e_{n,\beta_i}(s)) - \frac{e_{n,\beta_i}(s)}{2}\right)\frac{1}{\alpha^n}ds \\
&= (1-\alpha^n)\beta_i + \alpha^n\left(\frac{1}{2}+\epsilon_{old}\right) + \frac{1}{2\alpha^n}\int_{(1-\alpha^n)\beta_i+\alpha^n\epsilon_{old}}^{(1-\alpha^n)\beta_i+\alpha^n(1+\epsilon_{old})} \left((1-\alpha^n)(1-\beta_i)+s+\epsilon_{old}\alpha^n\right)^2 ds \\
&= (1-\alpha^n)\beta_i + \alpha^n\left(\frac{1}{2}+\epsilon_{old}\right) + \frac{1}{6\alpha^n}\left(1-(1-\alpha^n)^3\right) \\
&= (1-\alpha^n)\beta_i + \alpha^n\left(\frac{1}{2}+\epsilon_{old}\right) + \frac{1}{6}\left(3-3\alpha^n+\alpha^{2n}\right) \\
&= (1-\alpha^n)\beta_i + \alpha^n\epsilon_{old} + \frac{1}{2} + \frac{\alpha^{2n}}{6}
\end{aligned}$$

A.2 Threshold values for α

Optimal retention : The decision to retain/exclude current driver is optimal when cutoffs values of β found in 14 and 15 are equal:

$$\begin{aligned}
\bar{\beta} = \bar{\beta}_{SP} &\Leftrightarrow \frac{\epsilon_{new} - \alpha_{opt}\epsilon_{old}}{1-\alpha_{opt}} + \frac{1}{6} + \frac{\alpha_{opt}}{6} = \frac{1}{2} + \epsilon_{new} \\
&\Leftrightarrow \alpha_{opt} = \frac{3+6(\epsilon_{new}-\epsilon_{old}) - \sqrt{(3+6(\epsilon_{new}-\epsilon_{old}))^2-8}}{2}
\end{aligned}$$

From this it follows that if $\epsilon_{new} > \epsilon_{old}$, there always exist $\alpha_{opt} \in [0,1]$ such that type retention by the principal corresponds to the one of a benevolent social planner.

Full exclusion : The condition for full exclusion of a population is that $\bar{\beta} = 1$. Define $\alpha_{exclusion}$ the solution to this equation:

$$\alpha_{exclusion} = 3 - \sqrt{4 + 6(\epsilon_{new} - \epsilon_{old})}$$

Note that as soon as $\epsilon_{new} > \epsilon_{old}$, there exist $\alpha_{exclusion} \in [0,1]$ such that if $\alpha > \alpha_{exclusion}$, all types from incumbent population will be dismissed.